

Dynamiske kohortmodeller:

Modellering av levetid i forsikring

av

Jasprit Singh Toor

MASTEROPPGAVE

i studieprogrammet

Modellering og Dataanalyse

med studieretning

Finans, forsikring og risiko



Veileder: Erik Bølviken

Mai 2015

Det matematisk-naturvitenskapelige fakultet
Universitetet i Oslo

Forord

Denne masteroppgaven ble utført og skrevet som en del av min mastergrad ved Universitetet i Oslo (UiO) i studieprogrammet Modellering og Dataanalyse, med retning Finans, forsikring og risiko. Arbeidet med oppgaven gikk over perioden august 2014 til mai 2015, og inneholder et omfang på 60 studiepoeng.

Aller først vil jeg gjerne takke og bukke for min professor og veileder Erik Bølviken for et utrolig spennende tema for masteroppgaven. Erik har bistått med oppmuntrende råd, hjelp og inspirasjon underveis i oppgaven, og har gitt lærerike tilbakemeldinger under hele perioden med arbeidet.

Jeg vil takke mine foreldre Rajinder Singh Toor og Baldev Kaur, min lillebror Satinderjit Singh Toor, som alle har konstant bidratt med kjærlighet og støtte fra dag en på mitt første semester på UiO, og særlig under arbeidet med oppgaven. Sist men ikke minst vil jeg takke alle gode venner og medstudenter som både har bidratt med oppmuntring og korrekturlesing, i tillegg til gode avbrekk under studiene.

Innhold

1	Innledning	3
1.1	Bakgrunn	3
1.2	Mål	3
1.3	Data og metoder	3
1.4	Innsikt	3
2	Dødelighetsbegreper og modeller	5
2.1	Kohort og kohortdødelighet	5
2.2	Matematisk formulering	5
2.3	Problemer med kohortdødeligheten	7
2.4	Tidligere kohortestimeringer	7
3	Innledende datastudier og analyse	9
3.1	Dataene	9
3.2	Yngre kohorter	10
3.3	Gompertz-Makeham og maksimum-likelihood	12
4	Gompertz-Makehams modell og residualmodellering	17
4.1	Estimering av parametere i Gompertz-Makeham	17
4.2	Periodedødelighet og kohorteffekter	21
4.3	Minste kvadraters metode og residualer	28
4.4	Dynamisk modell og simulert utvikling	31
5	Pensjonsberegninger	38
5.1	Innledning og antakelser	38
5.2	Beregningsmetode	40
5.3	Resultater	42
6	Konklusjon	45
6.1	Svakheter med modellen	46
6.2	Fremtidig arbeid	47
7	Appendix	48
8	Litteratur og kilder	75

1 Innledning

1.1 Bakgrunn

Dødelighetssannsynligheter endrer seg over tid. Erfaringen i Norge og andre vestlige land tilsier omtrent ett års økning per tiår i forventet levealder slik denne størrelsen måles gjennom tradisjonelle periodedødeligheter. Dette er fenomener av stor betydning for befolkningsfremskrivninger og for fondering av pensjonsporteføljer i forsikring. Gode stokastiske modeller trengs for å bedømme risikoen til dem som har det økonomiske ansvaret for slike porteføljer der man de siste par tiårene har sett og utstedt såkalte dødelighetsobligasjoner som overfører deler av risikoen til en re-forsikrer. Spørsmålet er hva en "fair" pris for slike kontrakter egentlig er.

1.2 Mål

De vanligste dødelighetsmodellene i bruk i forsikring er basert på periodedødeligheter fra ulike år. Disse stammer fra ulike fødselsår (kohorter), informasjon som ofte neglisjeres. Denne oppgaven skal lage modeller ut fra de enkelte kohortene. Parametriske formuleringer må til da tidligere modeller kanskje ikke passer til våre tilfeller. Variasjonen mellom ulike år skal studeres for så å komme frem til enkle dynamiske formuleringer. Arbeidet og metodene som blir til, skal gjøres for norske kvinner og menn, men må også kunne gjentas for andre land, for eksempel i Skandinavia. Det er i tillegg også et mål å komme frem til et opplegg for fremskrivninger under slike kohort-baserte modeller.

1.3 Data og metoder

Jeg vil bruke dødelighetsdata for Norge lastet ned fra databasen [Human Mortality Database](#) (HMD-databasen). I tillegg vil dødelighetsdata for både Sverige og Danmark bli brukt, da modellene jeg kommer frem til i oppgaven skal prøves på disse. Programvaren som vil bli brukt i oppgaven er **R**.

1.4 Innsikt

Ser man på de siste 14 årene, kan man se at både menn og kvinner har økt sin forventet levealder. I denne perioden har menn økt sin forventede levealder med 2.51 år, til 74.84 år. Tilsvarende økning for kvinner har vært 1.45 år, til 80.61 år ¹. Dødelighetsnedgangen for menn skyldes hovedsaklig redusert hjerte-og kardødelighet, men generelt har norsk helsevesen (forsovet helsevesen i hele Vesten) blitt betydeligere bedre, og folk lever lenger. Hvilken betydning har det at folk lever lenger?

Kun et års økning i forventet levealder påvirker en pensjonsportefølje. Som skrevet ovenfor har vi i de siste 14 årene fått en økning på ca 2 års forventet levealder (grovt beregnet for både menn og kvinner). Man kan bare tenke seg hvor mye påvirkning det

¹Tall hentet fra rapport 96/6 av Mamelund og Borgan

vil ha på en portefølje dersom forventet levealder fortsetter å stige, altså vil effekten av lengre liv ha mye å si i forsikrings-og pensjonsverden. Som beskrevet i "Mål for oppgaven", baserer de vanligste dødelighetsmodellene seg på periodedødeligheter fra ulike år. Jeg skal lage modeller utifra de enkelte kohortene, og om de modellene viser seg å være optimale eller brukbare, så vil disse kunne brukes til nemlig å beregne forsikrings-og pensjonsporteføljer, noe som også er tilfellet i denne oppgaven. Ergo vil det ha lite mening å se på personer med alder lavere enn 18, så i denne oppgaven vil det bli fokusert på personer med alder og kohorter mellom 18-100.

2 Dødelighetsbegreper og modeller

2.1 Kohort og kohortdødelighet

Periodedødeligheten, som de vanligste dødelighetsmodellene baserer seg på, er basert på dødsfallene i en begrenset tidsperiode, gjerne et kalenderår, uansett når personene er født. Kohortdødeligheten derimot er basert på dødsfallene for personer født i en begrenset tidsperiode, altså ett eller flere fødselskull, uansett når personene dør. Det vil si, hvis vi ser på et fødselskull (om det enten er personer født i 1910, eller om vi velger en periode 1910-1914), så vil kohortdødeligheten basere seg på dødsfallene til disse menneskene. Hvis vi i det dynamiske tilfellet betegner periodedødeligheten lik q_{xk} , så er dødeligheten i kohort k lik:

$$\tilde{q}_{xk} = q_{xk-x},$$

der x er alder.

§ Definisjon 1. *I statistikk, så er en kohort en gruppe av individer eller subjekter, som har delt en bestemt hendelse sammen i et bestemt tidsrom.*

Eksempler på dette kan være folk som er født i Europa mellom 1918 til 1939, eller det som kommer til å være vårt tilfelle, kull som er født i samme årstall. Fra kohortene kan vi beregne kohortdødelighetene. Men la oss se fort tilbake på seksjon 1.2; ”De vanligste dødelighetsmodellene i bruk idag er basert på periodedødeligheter fra ulike år. Disse stammer fra ulike fødselsår (kohorter), informasjon som ofte neglisjeres.” Så hvorfor ikke fortsette med å bare bruke periodedødeligheter? Hva er forskjellen på periodedødelighet og kohortdødelighet?

§ Definisjon 2. *Kohortdødelighet er dødelighetsraten for en bestemt kohort.*

En overlevelsestabell med levealder for en *periode* gir et tverrsnitt av dødeligheten. Det er vanligvis slike data som presenteres når vi taler om levealder eller forventet levealder. Forventet levealder for en periode blir altså beregnet på grunnlag av dødeligheten for personer i en rekke generasjoner, og som kan ha vokst opp under svært ulike kår. Resultatene kan derfor være påvirket av de forskjellige generasjonenes levekår tidligere i livet. Men dersom vi studerer overlevelsestabeller for *kohorter* (fødselskull) fra fødselen og frem til hele kohorten er dødd ut, finner vi ut hvor lenge personer i en kohort levde i gjennomsnitt, og unngår dermed problemet med at personer født til ulike tider er vokst opp under ulike leveforhold.

2.2 Matematisk formulering

Stokastisk modellering i livsforsikring begynner med *hvor* lenge mennesker lever. Aktuarielle evalueringer på et tidsintervall T teller livslengden som $Y = LT$, der vi har at T kan være i år, kvartaler eller måneder og L er en parameter. Sannsynlighetsfordelingen til L er kjent som overlevelsessannsynligheter, og er som regel spesifisert gjennom de betingede sannsynlighetene:

$$\underbrace{{}_k p_l = \Pr(L \geq l + k | L \geq l)}_{\text{overlevelsessannsynligheter}} \text{ og } \underbrace{{}_k q_l = \Pr(k + l - 1 \leq L < l + k | L \geq l)}_{\text{dødelighetssannsynligheter}}. \quad (1)$$

På venstre side er overlevelsessannsynlighetene "likelihooden" av å leve k perioder lenger mens til høyre er dødelighetssannsynligheten sannsynligheten for at et individ dør i løpet av siste periode. Som man kan se av ligningene, avhenger begge av foreløpig alder l .

Det vi vil ha er 1-steps dødelighetsraten ${}_1 q_l = q_l = 1 - p_l$. Fra dette får vi rekursjonen:

$${}_{k+1} q_l = q_{l+k} \cdot {}_k p_l, \quad k = 0, 1, \dots \quad (2)$$

Systematiske feil forekommer når mennesker nå lever lenger og lenger, både grunnet forbedret helsevesen og at mennesker generelt liker å ta bedre vare på kroppen sin. Men som skrevet i avsnitt 1.4 er dette også av finansiell bekymring. Fra elementær statistikk har vi at dødelighetene er gitt ved en sekvens q_{l0} . Forlenger vi denne sekvensen til å være tidsavhengig, så har vi at q_l i alder l vil være avhengig av fødselsår, noe som vil være tilfelle for kohort-tilnærming. Kohorter født i $T_{-i} = -iT$ er da tilegnet dødeligheter $q_l = q_l(i)$. En enkel modell for dødeligheter av kohorter vil da være:

$$q_l(i) = q_{l0} e^{-\gamma(i)}, \quad (3)$$

der vi har at $\gamma(i)$ er en parameter som gjør at dødelighetsratene avviker fra standard sekvens q_{l0} , og tilnærmer den for kohorter. En positiv $\gamma(i)$ betyr at fødselskull fra kohort i lever lenger enn først antatt. I praksis så vil $\gamma(i) = \gamma_l(i)$, der vi også avhenger modellen av alder, noe som også fremstår i Lee-Carter modellen. Med dette, så har vi at rekursjonen fra ligning (2) nå vil bli:

$${}_{k+1} q_l = q_{l+k}(l) \cdot {}_k p_l, \quad (4)$$

for $k = 0, 1, 2, \dots$, med startverdi lik ${}_0 q_l = 1$. Begrunnelse: Vi er i tid $T_0 = 0$, og mennesker i alder l er født i $-l$ (og tilhører kohort l). Ved å sette inn ligning (3) inn i ligning (4), får vi sluttresultatet:

$${}_{k+1} q_l = q_{l0} e^{-\gamma(l)} \cdot {}_k p_l, \quad (5)$$

for $k = 0, 1, \dots$, og alle dødelighetene for ${}_kq_l$ har blitt beregnet.

Dette er da den matematiske formuleringen for beregninger av dødelighetsrater for kohorter. Vi skal i kapittel 3 og 4 se hvordan disse blir flettet inn i modelleringene våres, og at både Lee-Carter og Gompertz-Makeham modellene kan være av nytte når vi skal modellere dødelighetsrater i henhold til kohorter, især Gompertz-Makeham som er en populær matematisk beskrivelse av nemlig dødeligheter.

2.3 Problemer med kohortdødeligheten

Et problem med kohortdødeligheten er at vi må vente over 100 år for å finne gjennomsnittlig levealder for en kohort. For det er først når alle i kohorten er død ut (la oss si at en person i kohorten utrolig nok ble 120 år, så må vi vente med å regne gjennomsnittlig levealder helt til denne personen faktisk har død) at vi kan beregne kohortdødeligheten. Periodedødeligheten blir derfor ofte brukt som et estimat for den framtidige kohortdødeligheten, og er også grunnen til at kohortinformasjon neglisjeres. Periodedødeligheten kan også oppfattes som kohortdødeligheten for en hypotetisk kohort. For kohorter der de fleste er døde, kan vi estimere forventet gjenstående leveår med periodedødeligheten. Og dersom dødelighetsutviklingen for de aller eldste ikke bedrer seg betraktelig, vil vi på denne måten få et godt estimat for den faktiske kohortdødeligheten. Tar vi med dataene våre i betraktning, så må vi også huske på at kohorter kort tilbake i tid (la oss si rundt 1950 og oppover) ikke er fullstendige. Jo yngre kohortene er, desto mer usikkert vil estimatet for kohortdødeligheten bli, og jo nærmere kommer vi periodedødeligheten.

2.4 Tidligere kohortestimeringer

Det eksisterer mange ulike tilnærminger til modellering av trender og usikkerhet i dødelighet. Storparten av disse tilnærmingene omfatter effekter på tvers av alder, og økende kohort-effekter har blitt mer anerkjent som har resultert i en **alder-periode-kohort** modeller. En banebrytende modell for å studere dødelighetstrender og prognoser ble utviklet av Lee-Carter (1992). Lee-Carter-modellen er en numerisk algoritme som brukes i dødelighetsprognoser og forventet levealder-prognoser. Men det er ikke selve basismodellen som er av interesse. Lee-Carter familien av modeller har blitt utvidet av mange forfattere for å ta opp flere funksjoner i dataene. Av disse funksjonene så er en kohort-effekt den hyppigste funksjonen brukt, selv om det er mange modeller for ulike aldersgrupper, kohort- og periode effekter.

Forfattere som har brukt en slik utvidet kohort-effekt for å teste beregninger gjort på kohortdødeligheten har alle utvidet sin egen funksjon for å tilpasse dataene;

1. - Renshaw & Haberman (2003/2006)
2. - Brouhns *et al* (2002)
3. - Lee (2000)
4. - Cairns *et al* (2006,2009)

for å nevne noen. Disse store forfatterene har alle forsket på og testet multiple modeller på kohort dødelighetsdata. Det naturlige spørsmålet som treffer en da er: Hvorfor ikke utvikle/utvide en modell som fungerer for alle dødelighetsdata? Et logisk svar kan være fordi alle data er forskjellige. I en forskning gjort i *North American Actuarial Journal*², gjennom estimeringer og beregninger der de bruker data for menn i England og Wales opp mot data for menn i USA, kom de frem til at:

”På grunnlag av Bayes Information Criterion (BIC), finner vi at for høyere aldersgrupper, så vil en forlengelse av Cairns-Blake-Dowd (CBD), en modell som inkorporer en kohorteffekt, passe data for England og Wales best, mens for USA-data, så vil Renshaw og Haberman (RH)-forlengelsen av Lee-Carter modellen som også tillater for en kohort-effekt passe best.”

Så det ble brukt to forskjellige modeller på to forskjellige datasett. Det nevnes også i forskningen at det oppstod problemer angående hvor robust parameterene var under RH-modelleringen, men at ved å utvide CBD modellen ytterligere, slik at den også tillater for kvadratisk alders-effekt, så fikk man mye mer stabilitet av parameterene, selv om denne var dårligere i henhold til BIC-kriteriet. Og når man også tenker over hvor overraskende få programvarepakker for prognoser og beregninger det finnes med Lee-Carter modellen, så må kanskje hvert datasett bli håndtert annerledes.

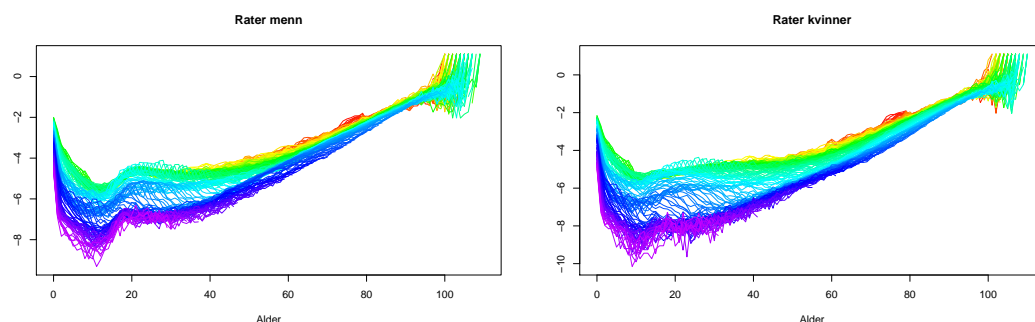
²North American Actuarial Journal, Volume 13, Issue 1, 2009

3 Innledende datastudier og analyse

Som oftest er det periodedødeligheten som brukes i anvendelser innenfor aktuarielle og forsikrings/pensjons- beregninger. Dette medfører at det er mange ulike programmer og modeller som estimeres over periodedødeligheten. Finnes det slike programmer og modeller for kohortdødeligheten?

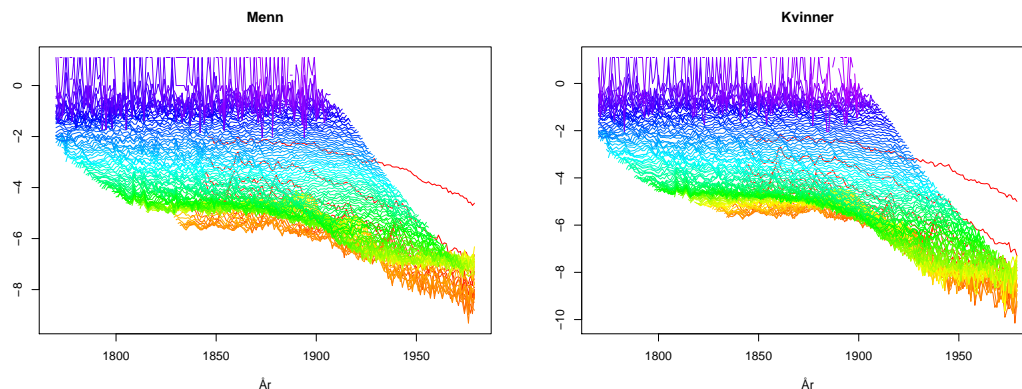
3.1 Dataene

Som nevnt tidligere henter vi datasettet vårt fra Human Mortality Database (HMD), der vi får dødeligheter for kohorter for den norske befolkningen. Datasettet er delt opp for både menn og kvinner. Vi skal gjennom oppgaven prøve ut kohortene for hvert årstall (1846,1847,...,1920,1921 osv), men også prøve å modellere for kohorter født i 5 års perioder (altså 1846-1850, 1851-1855,...), for tilslutt å se på kohorter i 10-års perioder for å se hvordan dødelighetene endrer seg for de forskjellige folkegruppene og kohortene. Første modelleringen går fra kohorter helt tilbake til 1800-tallet, og opp til midten av 1900-tallet. De påfølgende figurer er for den norske befolkningen, der vi plotter den logaritmiske dødelighets- sannsynligheten mot alder og år, for både menn og kvinner, respektivt.



Figur 3.1: Dødelighetsprosjeksjon for alle aldre

Vi ser fra den logaritmiske dødelighetssannsynligheten at den øker jevnt med alderen. Fra 50 år og oppover øker den proporsjonalt med alderen. Det er interessant å nevne at før fylte 50 år, så er det både opp- og nedgang i dødelighetssannsynligheten. Fargene indikerer de forskjellige dødelighetsratene i henhold til alderen i de respektive årstallene, og fargene som strekker seg lengst er de eldre kohortene som har fullstendige dødelighetsrater for alle aldre. I forrige århundre var det stor barnedødelighet blant befolkningen, ergo den høye raten for alder lik 0.



Figur 3.2: Dødelighetsprosjeksjon gitt mot år

De norske dataene bekrefter at dødeligheten er fallende i alle aldre med forskjellig forløp i henhold til de ulike aldre. Igjen så indikerer fargene dødelighetsratene for de ulike årene.

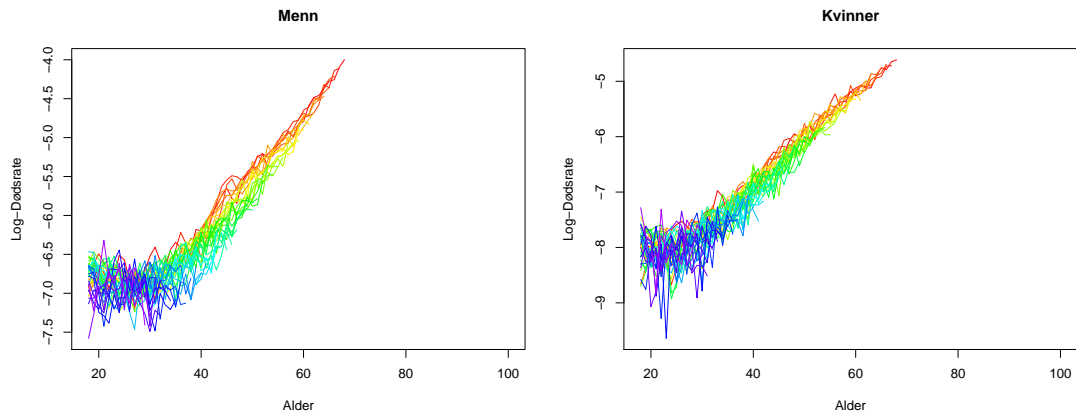
Det er interessant å se hvordan kohortdødeligheten var for de fullstendige kohortene så langt bak i tid, men for oss som lever i 2015 så er det av lite betydning. Vi vil se på kohortene nærmere vårt århundre da det er disse som skal modelleres videre i for eksempel forsikrings- og pensjonstilfeller. La oss si at vi begynner med å se på kohortene fra 1940 tallet og oppover, da vi både har at 2 verdenskrig spilte en rolle i de ulike kohortenes dødelighetssannsynligheter, og at mange av kohortene nå ikke er fullstendige. Jo yngre kohortene blir, altså jo nærmere vi kommer vårt inneværende årstall, jo mer ufullstendige vil de være (avsnitt 2.2). Men så er det av interesse å se om disse ufullstendige kohortdødelighetene kan brukes i statistiske beregninger videre.

3.2 Yngre kohorter

Forskjellen for yngre kohorter og de vi så på i forrige tilfelle (kohortene fra 1846 og oppover) er alle de forskjellige sykdommene som florerte og tok mange menneskeliv. I perioden 1850-1920 herjet mange smittsomme sykdommer som f.eks: kolera, flekktyfus, difteri, skarlangensfeber, tyfoidfeber, meslinger, tuberkulose og influensa (spanskesyken). Disse sykdommene tok mange liv, og spesielt var barnedødeligheten veldig høy. Dødeligheten generelt var høy for de fleste av kohortene.

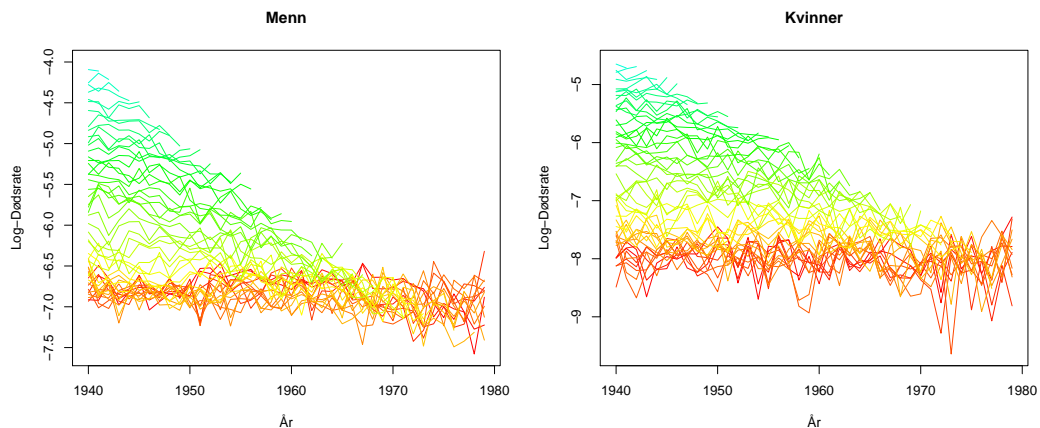
Kohortene fra 1940-tallet og oppover er preget av mye mindre død forårsaket av sykdommer. Fra 1930-tallet var det store fremskritt innenfor det medisinske området i kampen mot sykdommer. Generelt er dødeligheten mye lavere grunnet sykdom, og folk begynner å leve lenger og lenger. Dette er et problem for kohortdødeligheten da den trenger at individer i hver kohort skal dø ut før vi får eksakt kohortdødelighet for den gitte kohorten. Videre ser vi hvordan dødeligheten endrer seg for både kvinner og menn i forhold til de foregående kohortene.

3.2 Yngre kohorter



Figur 3.3: Dødelighetsrater for kohorter fra 1940 og oppover, *alder*

Sammenligner vi disse dødelighetene mot den forrige (Figur 3.1) ser vi med en gang at dødelighetene for disse kohortene er mye lavere. Fra fylte 40 år og oppover kan vi si at dødeligheten så å si proporsjonalt øker med alderen, men før fylte 40 så fikk noen av kohortene et lite hopp. Dette kan være forårsaket av at det var krig og at noen av sykdommene ikke var kurerbare på denne tiden. Ser man på forskjellene mellom menn og kvinner så fremstår det av figuren at dødeligheten for menn er litt høyere enn den for kvinner, noe som kan fremkomme av grunnene nettopp beskrevet. Vi må også regne med at mange av disse kohortene har individer som fortsatt lever den dag i dag (et individ født i 1970 vil ikke ha kohortdødelighet for alder opptil 100, da dette ville resultert i at vi måtte kunne se inn i fremtiden til år 2070). Med andre ord at vi har ufullstendige kohorter. Det er også derfor figuren tynnes kraftig ut i halen, fordi det ikke er data for de eldste aldre.



Figur 3.4: Dødelighetsrater for kohorter 1940 og oppover, *år*

Ser vi nå på forskjellene mellom kvinner og menn i årstall så begrunner også denne figuren argumentet vårt fra tilsvarende forrige figur, nemlig at kvinner innehar lavere dødelighet enn menn når man ser på dødelighet for de forskjellige kohortene med henhold til både alder og år. Spesielt fra denne figuren ser vi de ulike dødelighetene til menn og kvinner, der kvinners dødelighet går i en såkalt skjev linje nedover, mens for menn er dødeligheten litt kurvet. Kurvene tynner seg ut i halen fordi vi har kohorter født nærmere vårt årstall, men som ikke har dødd ut ennå (18-åring kohort fra 1979 vil idag være 54 år, og i dagens samfunn er det veldig få 54-åringer som dør). Sammenligner vi denne figuren mot figur 3.2 ser vi at det nå er mye mindre data som modelleres.

Det ble forklart innledningsvis at det også skulle prøves ut beregninger for både 5 og 10-års kohorter. Dette ble modellert på og testet, men når vi allerede har minimalt med beregninger for de yngre kohortene så finnes det rett og slett ikke nok data for disse tilfellene til å fremstille gode resultater.

Oppsumert har vi sett at uansett periode og år, så har kvinner hatt lavere dødelighet enn menn på alle alderstrinn. Dødeligheten i forrige århundre var aldeles større enn det den er i vårt århundre, noe som kommer av store medisinske fremskritt og mindre verdenskriger. Kohortene og kurvene estimert blir tynnere jo nærmere vi kommer vår egen levetid, noe som er grunnet ufullstendige kohorter. Det vil videre i oppgaven kun fokuseres på kohorter fra år til år, altså ikke 5 og 10-års kohorter, da disse inneholder mangelfull data.

3.3 Gompertz-Makeham og maksimum-likelihood

Det er allerede nevnt i avsnitt 2.2 at den mest populære matematiske beskrivelsen av dødeligheter går helt tilbake til 1860, ved navnene Gompertz og Makeham, som begge hadde separate bidrag. Det finnes flere artikler og research-studier der Gompertz-Makeham blir brukt på kohort data, og derfor skal det i denne oppgaven også prøves ut modeller av denne typen på våre kohort data. Intensiteten er nå en parametrisert kurve av formen:

$$\mu(x) = \theta_0 + \theta_1 e^{\theta_2 x}, \quad (6)$$

der θ_0 , θ_1 , og θ_2 er parametre.

Dette er et resultat vi bruker gjennom videre beregninger. I praksis er dødsintensiteter tilnærmet lik dødssannsynligheter, det vil si: $\mu_x(l) \approx q_l$. Vi kan da anta at

$$q_l = \theta_0 + \theta_1 e^{\theta_2 l},$$

der q_l er sannsynligheten for at en person i alder l skal dø. Men vi skal ikke ha sannsynligheten for at en person i alder l dør, men sannsynligheten for at en person i kohort k med alder x dør. Fra avsnitt 2.1 har vi at dødeligheten i en kohort er gitt ved q_{xk-x} , slik at vi da får:

$$q_{xk-x} = \theta_0 + \theta_1 e^{\theta_2(xk-x)}, \quad (7)$$

der x er alder, og k er kohort. Gitt at n_l er lik antall individer som er utsatt for risiko for å dø i en viss kohort, i alder l , og at x_l er antallet blant dem som dør, også i alder l , så har vi at sannsynligheten for x_l er gitt ved

$$Pr(x_l) = \binom{n_l}{x_l} q_{xk-x}^{x_l} (1 - q_{xk-x})^{n_l - x_l},$$

som da er binomisk fordelt. For videre å estimere parameterene θ_0 , θ_1 og θ_2 må det brukes maksimum-likelihoodteori. Fra *Erik Bølviken* sin bok ³ så utledes det fra $\mathcal{L}(\theta_0, \theta_1, \theta_2) = \Pi Pr(x_l)$ et kriteriet for maksimum-likelihood estimatet:

$$\mathcal{L}(\theta_0, \theta_1, \theta_2) = \sum_l \{x_l \log(q_{xk-x}) + (n_l - x_l) \log(1 - q_{xk-x})\}, \quad (8)$$

der q_{xk-x} , gitt ved ligning 7, settes inn i kriteriet ovenfor og uttrykket vil da bli maksimert.

Vi skal nå bruke disse uttrykkene for våre kohortdata. Koden brukt er tilpasset slik at det passer kohortdataene. Dataene våre strekker seg fra 1940-1979 (yngre kohorter enn dette inneholder minimalt med data), og vi skal se på et eksempel der vi først minimerer minus-loglikelihood funksjonen. Kalkulus og tidligere matematiske beregninger forteller oss at å minimere minus-loglikelihood funksjonen er ekvivalent med å maksimere likelihood-funksjonen.

Vi begynner med å bruke faste tall for θ . I Appendix C kan all koding av resultater sees hvordan blir modellert.

```

1 minusgompakll1 = function(s, ll, n, z) {
2   t = exp(s)
3   p = exp(-t[1] - t[2] * exp(t[3] * ll))
4   -sum((n-z) * log(p) + z * log(1-p))
5 }

```

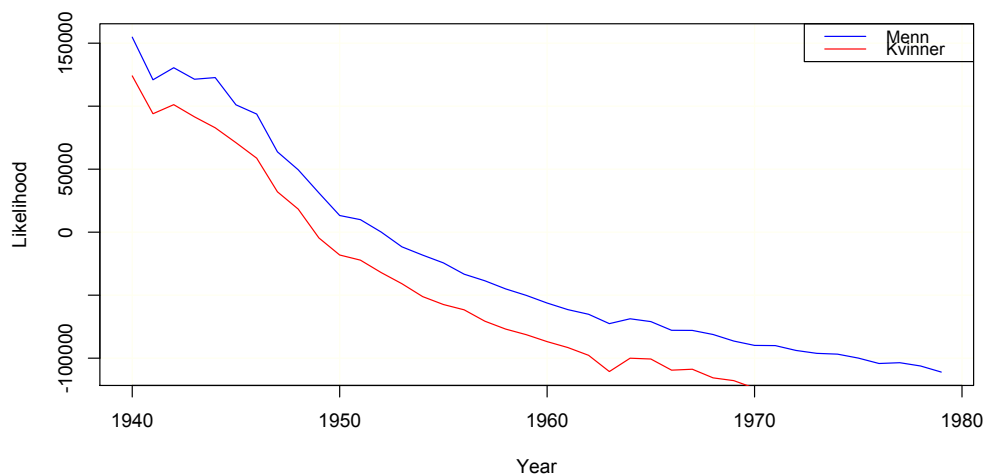
Vi har at s er θ -parameterene for kohorten, ll er en vektor med aldre 18-100, n er en vektor med antall individer utsatt for risiko for å dø, mens z er en vektor med antall døde individer for hver alder. Dette kjøres over alle år. Med tanke på hvor mye data det er, så har vi en tabell under med utvalgte estimater for diverse år, mens hele tabellen kan sees i Appendix B. Vi får:

³Computation and Modelling in Insurance and Finance. Cambridge University Press, Cambridge

År	Menn	Kvinner
1940	154782.34	124051.08
1950	13230.12	-18111.88
1960	-56237.43	-86865.35
1970	-89861.69	-123823.66
1979	-111014.94	-165795.31

Tabell 3.1: *Negativ log-likelihood estimer for utvalgte år, menn og kvinner*

Her har vi kjørt funksjonen `minusgompakll1` over kohortdataene våre. Funksjonen tar som sagt inn vektorer med alder, individer med risiko og antall døde, respektivt for både menn og kvinner. Disse estimatene har i og for seg lite mening med dem, men kan bli brukt til sammenligning med andre modellens estimer. Vi legger også merke til at estimatene er økende (vi beregner minus log-likelihood estimer, og jo mer negative jo bedre) for hver kohort opp til sist endte år, noe som kommer av mer og mer manglende data for nyere kohorter. Dette betyr at funksjonen passer bra til kohortdataene. Det som også er av interesse er nemlig figuren som kommer av programkodingen. Under er det plottet menn mot kvinner, deres log-likelihood mot år, respektivt:

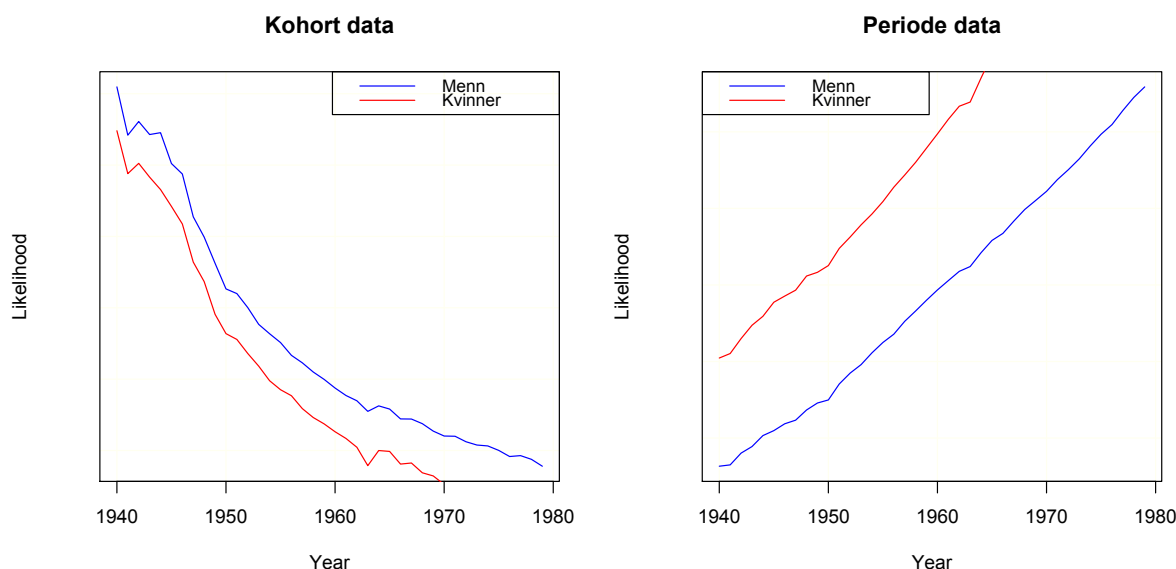


Figur 3.7: *Negativ log-likelihood plot for menn og kvinner*

Her ser vi likelihood estimatene for hver kohort for hvert år for menn og kvinner. Logisk har kvinne-kohortene litt lavere likelihood-estimer, som kommer av at menn i større grad har høyere dødelighetsrisiko. Det kommer også av at det var mindre antall dødsfall i kohortene for kvinner enn det var for menn, og dette skillet ser ut til å gå gjennom alle kohortene for alle år. Figuren gir en indikasjon på hvordan kohortene utvikler

seg. Dette tilsier at dødeligheten minker med årene. Rundt 1950 ser skillet mellom menn og kvinner ut til å være minst. Det kan komme av at andre verdenskrig nettopp ble avsluttet, og det førte til høyere antall fødte individer i kohortene, mens skillet ser ut til å være størst opp mot 1970 og oppover som igjen kan være av grunner som den kalde krigen. Da ville kohorter for menn, født i nettopp rundt 1945/50 være ca 20-30 år, og deres liv gikk tapt rundt denne perioden. Ergo vil risikoen for dødelighet for menn være betraktelig høyere enn for kvinners, og antall døde i kohorter for menn være flere.

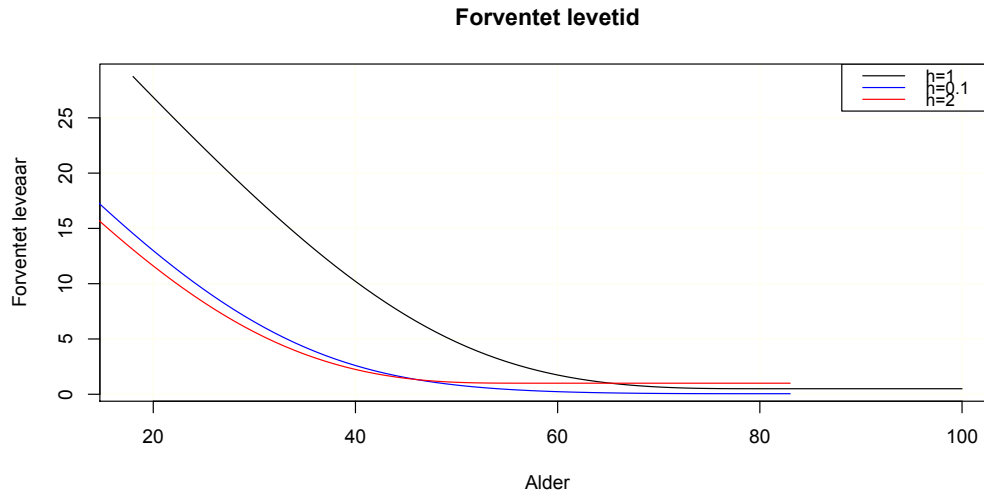
Av interesse kan vi også ta med figur av periodedataene. Fra kohortdataene ser vi på kohort-dødeligheten, mens i periodedataene er periodedødeligheten i fokus. Vi får:



Figur 3.8: *Kohort vs. Periode-data, for menn og kvinner*

I figuren til venstre ble det forklart ovenfor at det her ble sett på negativ log-likelihood for en gitt kohort, og at denne minket med årene. I figuren til høyre er det periodedødeligheten modellen blir estimert over. Men denne figuren gir ingen mening, for linjene her går oppover, noe som betyr at den negative log-likelihooden øker. Dette kommer av at modellen tilpasset er ment for kohortdata, og ikke periodedødeligheter, og at vi er på riktig vei.

Som nevnt tidligere er periodedødeligheten basert på dødsfallene i en begrenset tidsperiode, gjerne et kalenderår, mens kohortdødeligheten derimot baserer seg på dødsfallene til personer født i en begrenset tidsperiode, gjerne fødselskull. La oss videre se på hva forventet gjenværende leveår for en viss kohort er. Det brukes igjen en parametrisert modell under Gompertz-Makeham med faste parametere.



Figur 3.9: Forventet gjenværende leveår for en viss alder i kohort. h -parameter angir tids-inkrement brukt i parametriseringen.

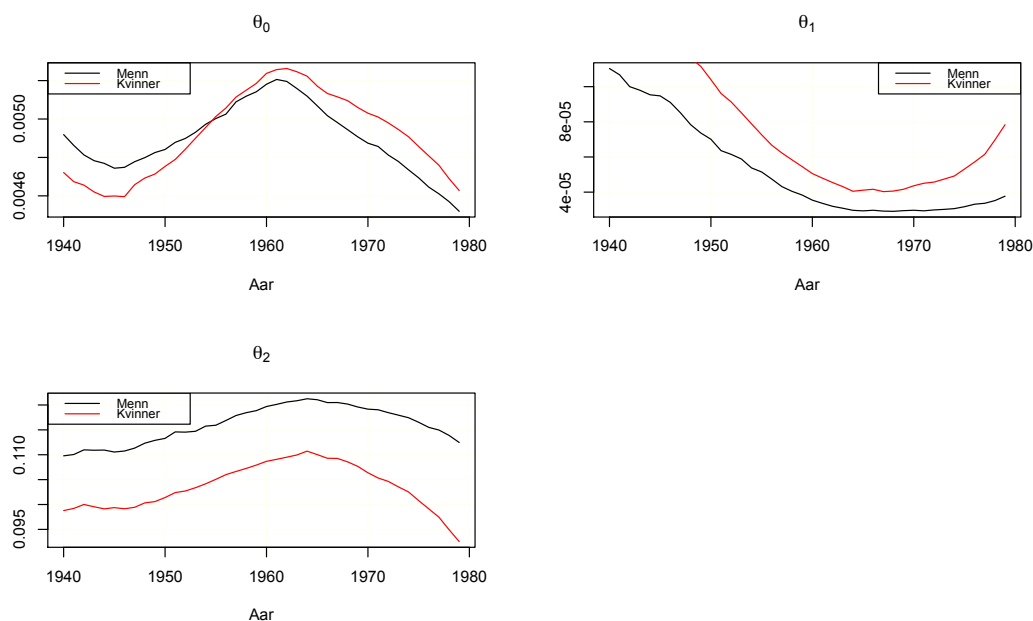
Tidsinkrement er noe som må bestemmes i funksjonen, og en h -parameter lik 1 betyr fra år til år. Vi ser med en gang forskjellen vi får i forventet levetid om tidsinkrement-parameteren endres fra 1 til $0.01/2$. La oss fokusere på tidsinkrement $h=1$ (regelmessig brukt). Ved første øyekast ser man at forventet gjenværende leveår for en 20-åring er ca 30 år. Logisk? En 20-åring vil i 1940 være født i kohort 1920. Forventet leveår minker proporsjonalt med alderen frem til 65, videre får vi såpass små estimerte gjenværende leveår. Forventet leveår for kohortene er lavere enn det man først forventet, og det er dessuten brukt faste parametere for alle kohortene slik at vi har en statisk modell, noe som ikke gir riktig bilde da det i realiteten skal være forskjellige parametere for hver kohort.

4 Gompertz-Makehams modell og residualmodellering

I forrige kapittel ble det brukt faste tall for θ -parameterene for både menn og kvinner. Spørsmålet er om disse er de mest optimale for modellen basert på kohortdataene. For å løse ligning (8) (maksimum-likelihood for parameterene) må det brukes innebygde numeriske metoder i R, `optim()`,⁴ og slik vil parameterverdiene θ_0 , θ_1 og θ_2 i Gompertz-Makehams modell estimeres, for kvinner og menn separat.

4.1 Estimering av parametre i Gompertz-Makeham

`optim()` er en metode for å estimere parameterverdier numerisk, enten ved å maksimere eller minimere. Vi bruker den standard innebygde quasi-Newton algoritmen BFGS. Metoden krever at vi legger inn startverdier for parametre som skal estimeres og vi bruker de faste parameterene fra siste avsnitt som startverdier, nemlig 0.000335, 0.000123 og 0.135335 henholdsvis for θ_0 , θ_1 og θ_2 , like startverdier for begge kjønn. Disse startverdiene er allerede brukbare (små), så metoden vil ikke iterere altfor mange ganger før den finner de optimale parameterene. Dette gjør vi så vi slipper vente altfor lenge med simuleringen.



Figur 4.1: *Parametre Gompertz-Makeham modell, for menn og kvinner*

⁴<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/optim.html>

Et vanlig problem med Gompertz-Makeham modellen er at den forlanger at sannsynlighetene for dødeligheter øker med alderen, altså ikke kan minke. Dette er et problem for oss, i og med at mange av kohortene ikke har fullstendige datasett, og dødelighets-sannsynligheter for yngre kohorter da vil mangle data. `optim()`-funksjonen krever at data vi sender inn i funksjonen er gitt ved et tall, altså ikke NA ("not a number"), som noen av kohort-dataene faktisk har som verdi. Dette problemet måtte omgås ved å sette opp en restriksjon for modellen vår:

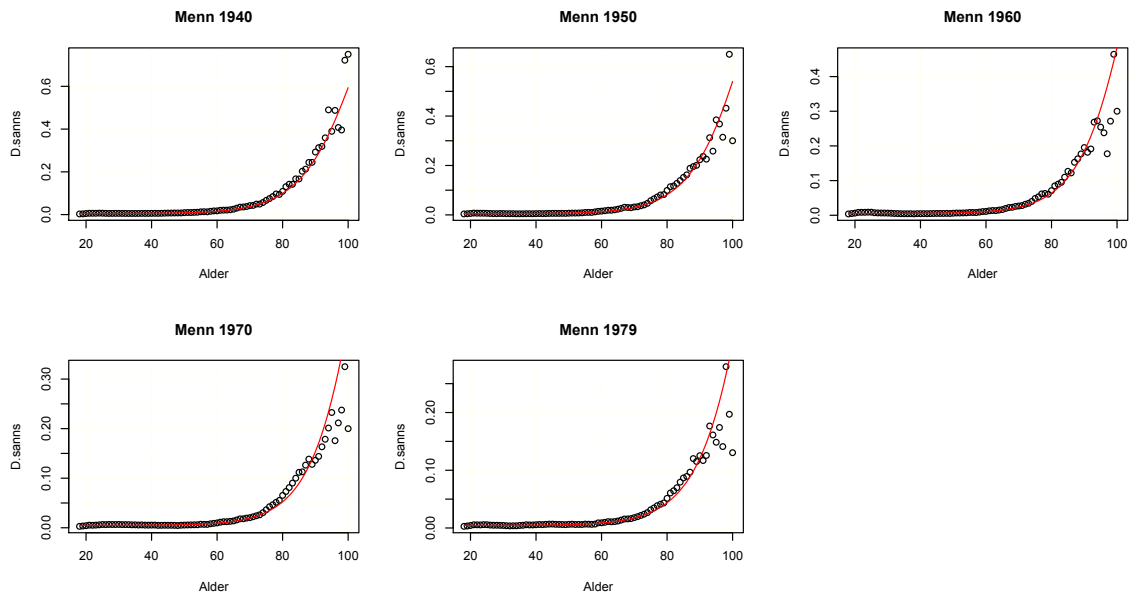
Restriksjon 1: $NA \rightarrow 0$.

Men dette kan føre til noen smårare parameter verdier, noe som kan sees av **Figur 4.1**, der kurvene framstiller parameter-estimatene. Videre vil vi også se at denne restriksjonen vil føre til en annen ulempe. Under er det lagt ved tabell over noen av estimatene for parameterene θ_0 , θ_1 og θ_2 , for utvalgte år.

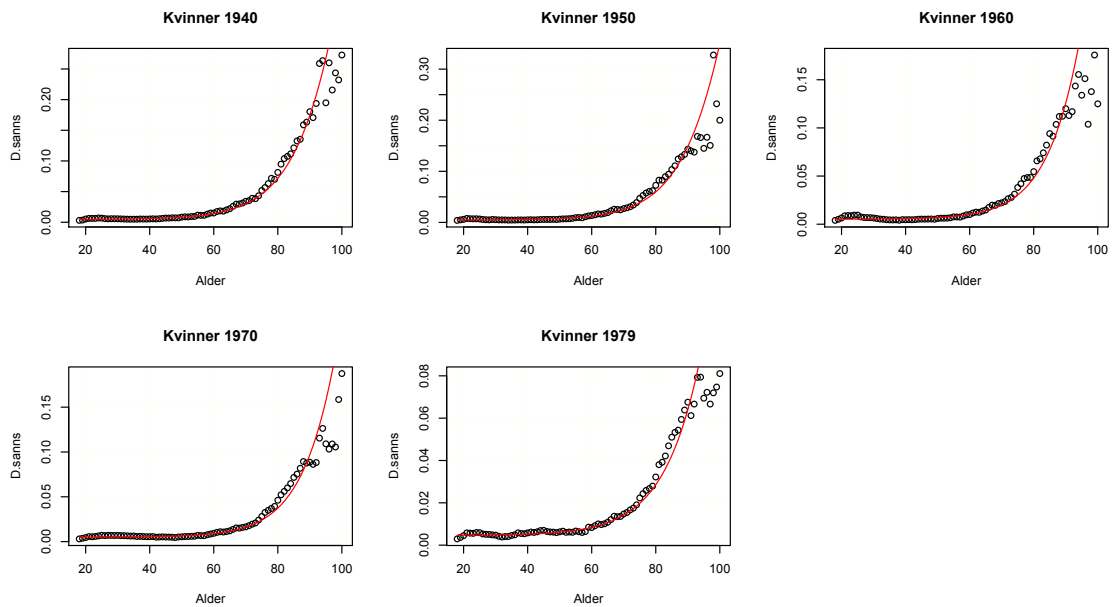
	Menn			Kvinner		
År	θ_0	θ_1	θ_2	θ_0	θ_1	θ_2
1940	0.0049	1.103365e-04	0.1097	0.0047	1.524423e-04	0.0987
1950	0.0048	7.008969e-05	0.1132	0.0047	1.039970e-04	0.1014
1960	0.0051	3.544324e-05	0.1196	0.0052	5.056321e-05	0.1086
1970	0.0048	2.968892e-05	0.1191	0.0050	4.362298e-05	0.1063
1979	0.0045	3.769288e-05	0.1124	0.0046	7.835645e-05	0.0925

Tabell 4.1: *Estimater for parametere θ_0 , θ_1 og θ_2 i Gompertz-Makeham modellen for utvalgte år, menn og kvinner*

4.1 Estimering av parametere i Gompertz-Makeham



Figur 4.2: Plott av rådata (dødelighetssannsynligheter) mot Gompertz-Makeham modell, *menn*

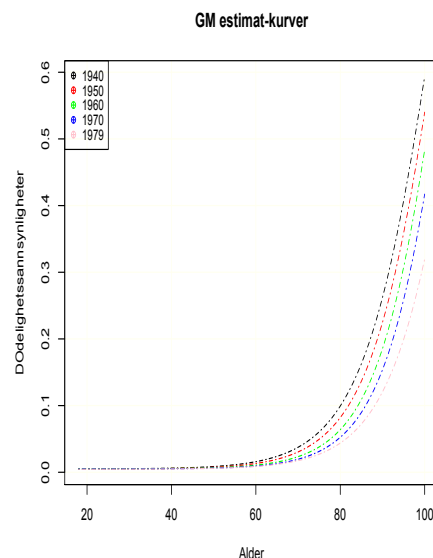


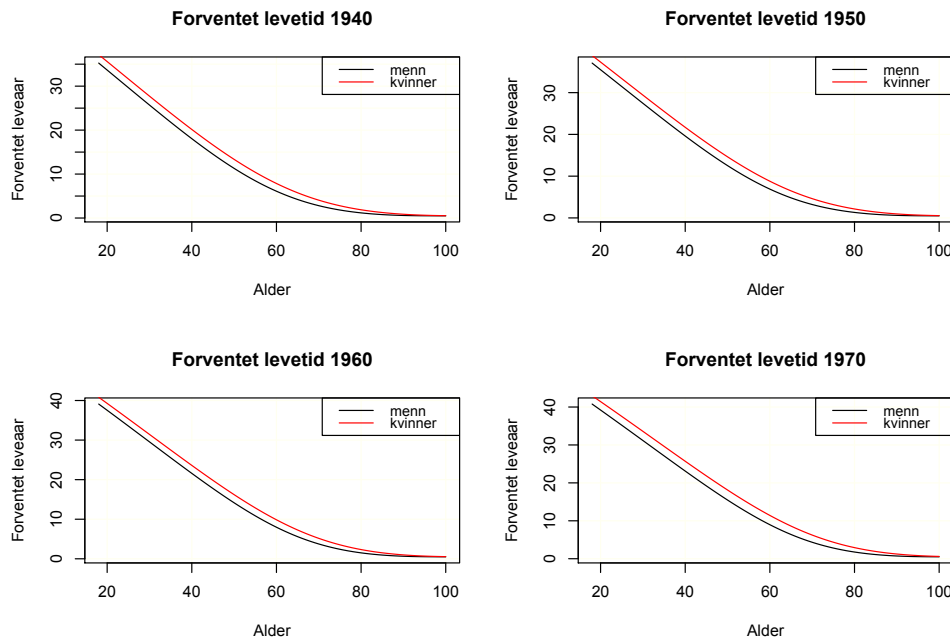
Figur 4.3: Plott av rådata (dødelighetssannsynligheter) mot Gompertz-Makeham modell, *kvinner*

Men til syvende og sist så er det tilpasningen av modellen opp mot rådataene som forteller oss hvor bra modellen passer til dataene. Ser vi på figur 4.2 og 4.3 så har vi dataene plottet mot Gompertz-Makehams modell, der de røde linjene er Gompertz-Makeham modell for hvert av årene vi simulerer over, for menn og kvinner respektivt. Gompertz-Makeham tilpasningen ser ganske lik ut for alle simuleringene, men legg merke til y-aksen i figurene (vertikal akse) der disse er tilpasset for hver estimering. Ut ifra resultatene ser modellen ut til å passe ganske bra opp mot rådataene, vi har med andre ord en bra modell. Vi ser at rådata-punktene tynner seg ut jo høyere opp i alder vi kommer, noe som skyldes av manglende data for visse aldre og oppover (faktisk mye ufullstendig for 60 år og oppover!). Derfor får vi også tunge haler for høye aldre. De eldre kohortene har mer data over alle aldre, og ut ifra disse estimerer modellen en bra tilpasning over aldersspekteret, med noe over og under-estimeringer (mest underestimert) i de tyngste halene.

For å få bedre overblikk over hvor ulike disse kurvene er i henhold til hverandre, har vi til høyre et plott over Gompertz-Makeham kurvene, for de forskjellige kohortene respektivt. Jo nærmere vi simulerer til inneværende årstall, jo mer synker kurvene. Tenker man på at de eldre kohortene inneholder mer fullstendige data over alle aldre, og de ikke er like fullstendige for yngre kohorter, så kan det se ut som modellen kanskje underestimerer dødelighetssannsynlighetene en smule for de mer ufullstendige kohortene. Dette kan komme av restriksjonen vi satte i `optim()`-funksjonen tidligere, der alle ufullstendige data ble omgjort til 0. I tillegg er dødeligheten for yngre kohorter også lavere enn de eldre.

Før vi går videre til sjekk og test av residualene, inkluderes det først en figur som estimerer forventet gjennværende leveår for individer over aldre, for utvalgte persentiler. I motsetning til figur 3.9 som brukte faste parametere for θ , så har vi nå et mye bedre oversiktlig plott som bruker parameterene estimert i tabeller Appendix B2 og B3, og vi ser hvor stor forskjell det er på modeller som bruker faste parametere i forhold til parametere som er stokastiske.





Figur 4.4: Gjennværende forventet levealder med tidsinkremerte parametere, for utvalgte kohorter, i år

Ved første øyekast ser disse estimatene veldig like ut. Kvinner har for alle kohorter, og alle aldre mer gjennværende levealder. Det som skiller seg ut er den vertikale aksene. For kohorter 1960/70 har vi mer levealder enn de foregående kohortene, noe som kommer av lavere dødelighet for nyere kohorter. For aldre over 100 år får vi 0 forventede levealder, m.a.o; for alle kohortene så vil alle individer ved inneværende alder 100, ha 0 år igjen å leve. Dette kommer av restriksjonen vi har satt gjennom hele oppgaven, at vi kun ser på individer med alder fra 18-100. Sammenlikner vi denne figuren mot figur 3.9, ser vi at vi får mye glattere kurver. Å estimere levetid med stokastiske parametere opp mot modellen, i motsetning til faste parametere, gir stor bedring av modellen.

4.2 Periodedødelighet og kohorteffekter

Vi har til nå tilpasset Gompertz-Makeham til kohortdataene våres, og sett på mange innledende studier på disse. Det som skal testes videre er hvorvidt de forskjellige resultatene er i henhold til data tilpasset til periodedødeligheten, og kanskje enda viktigere:

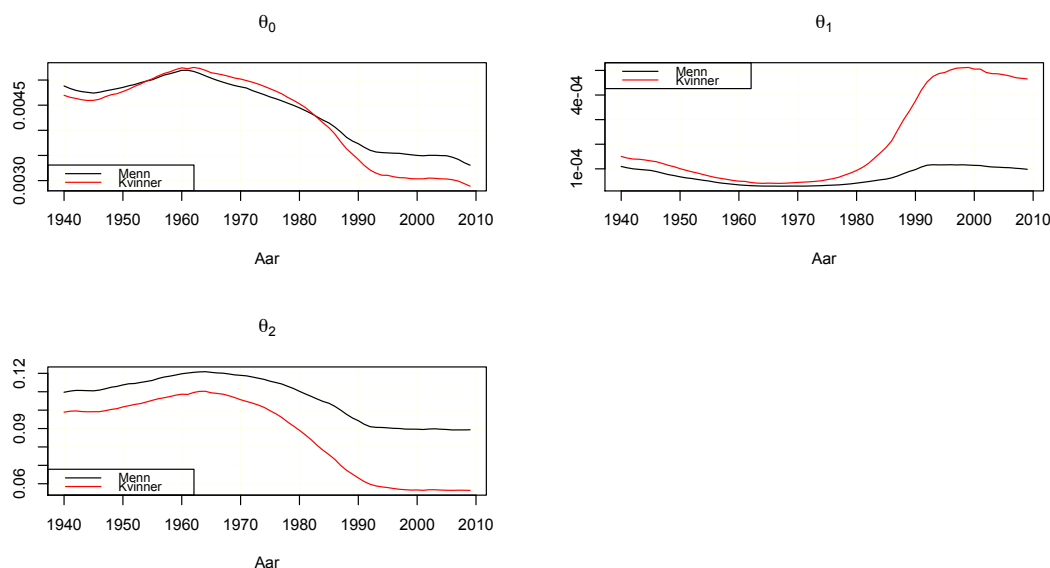
Hvor viktige er kohorteffektene når de utelates i periodedødelighetsmodeller?

Med dette mener vi om resultatene kan bli ytterligere bedre om vi ikke neglisjerer kohorteffektene inn i periodedødeligheten, eller om de ikke spiller noen rolle fordi de er ubetydelige. Vi skal teste om vi ser noen sammenheng eller kobling i residualene for

modellene våre. Ved ligning 7 har vi modellen for kohortdataene, og en mulig modell for periodedødeligheten kan være som følger:

$$\log \frac{q_x}{1 - q_x} = \theta_{0x} + \theta_{1x} e^{\theta_{2x} x} + c_k,$$

hvor x er alder og c_k er et kohortbidrag til periodedødelighetsmodellen. Før vi setter oss inn i dette, oppgir vi først resultater for periodedødeligheten, og sammenlikner disse med kohortdødeligheten. Husk at dataene fremstilt og modellert på nå gjelder for periodedødeligheten, og at vi som en fordel har fullstendige data. Modelleringen går nå også over individer utsatt for risiko for å dø, akkurat som i kohort tilfellet. Nevner igjen at periodedødeligheten er basert på dødsfallene i en begrenset tidsperiode, gjerne et kalenderår, uansett når personene er født, mens kohortdødeligheten derimot er basert på dødsfallene for personer født i en begrenset tidsperiode, altså ett eller flere fødselskull, uansett når personene dør.



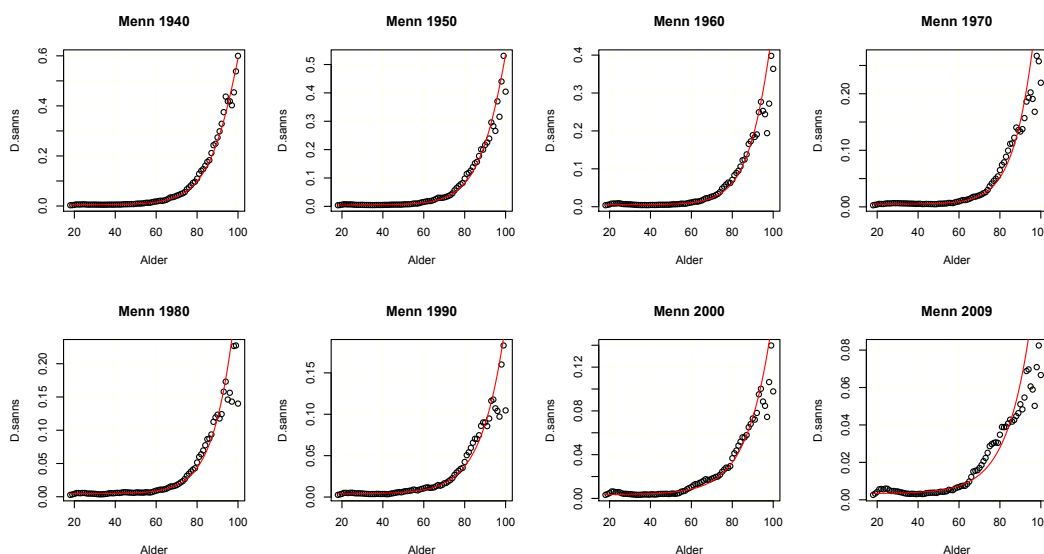
Figur 4.5: *Parametere for Gompertz-Makeham modell, for kvinner og menn, periode-data*

Vi har tilpasset Gompertz-Makeham modellen til periode-dataene våre, og estimert frem θ -parameterene. Sammenlikner vi denne figuren med Figur 4.1 oppfører kurvene seg helt annerledes. Dette er naturlig når vi nå modellerer over to forskjellige datasett, med helt forskjellige estimer for parameterene. θ -verdiene går fra år 1940 og helt opp til 2009, noe som igjen kommer av at vi nå har fullstendige datasett, og som er hovedgrunnen til at periodedødelighet foretrekkes fremfor kohortdødelighet.

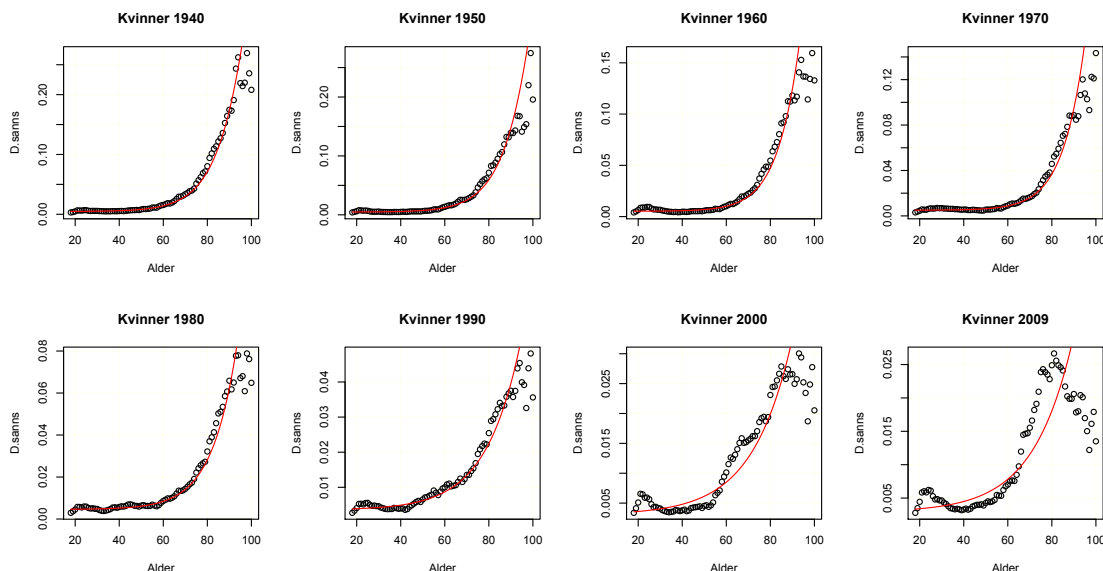
	Menn			Kvinner		
År	θ_0	θ_1	θ_2	θ_0	θ_1	θ_2
1940	0.0048	1.09e-04	0.1097	0.0046	1.50e-04	0.0988
1950	0.0048	7.25e-05	0.1129	0.0047	1.08e-04	0.1008
1960	5.15e-03	3.72e-05	1.19e-01	5.21e-03	5.28e-05	1.08e-01
1970	4.88e-03	2.99e-05	1.19e-01	5.03e-03	4.31e-05	1.06e-01
1980	4.49e-03	3.89e-05	1.11e-01	4.59e-03	8.31e-05	9.13e-02
1990	3.78e-03	8.97e-05	9.57e-02	0.0035	0.0003	0.0652
2000	0.0035	0.0001	0.0896	0.0030	0.0005	0.0565
2009	3.30e-03	9.81e-05	8.93e-02	0.0028	0.0004	0.0563

Tabell 4.2: Estimer for parametre θ_0 , θ_1 og θ_2 i Gompertz-Makeham modellen for utvalgte år, menn og kvinner, *periode-data*

I forrige avsnitt testet vi om modellen vi tilpasset kohortdataene faktisk passet rådataene (Figur 4.2 og 4.3). Denne prosedyren må vi også foreta i dette tilfellet, altså sjekke om disse estimatene for θ -parameterene er optimale. Vi må altså ha en modell-sjekk, om modellen tilpasset passer periodedataene.



Figur 4.6: Modellsjekk, rådata vs θ -verdier, menn, *periode-data*



Figur 4.7: Modellsjekk, rådata vs θ -verdier, kvinner, *periode-data*

Av begge figurene kan vi se at modellen passer veldig bra. Selv for periodemodellering har vi at data for årene nærme vårt inneværende årstall (2015), tynnes ut. Dette gjelder for begge kjønn, men vi ser at tilpasningen er bedre for menn enn kvinner. Dette kommer av at det er flere individer av menn som er utsatt for risiko for å dø, for alle aldre enn kvinner. Det trengs heller ingen restriksjon i dette tilfellet.

I ”Mål for oppgaven” (avsnitt 1.2) ble det forklart at dagens dødelighetsmodeller baserer seg på periodedødeligheter fra ulike år, som igjen stammer fra ulike fødselsår (kohorter), informasjon som ofte neglisjeres. Vi har nå fått satt opp en modell for både periode- og kohortdødeligheten, og det er blitt testet at dette er gode modeller. Innledningsvis til dette avsnittet ble det forklart hvordan vi skal se etter kohorteffekter og mønstre i residualene for periodedødeligheten. Vi skal se etter kohorteffekter i residualene for periodedødeligheten, og estimere størrelsen på denne effekten. Om det er stor effekt eller ingen, er begge interessante resultater, og vi skal teste om Gompertz-Makeham modellen klarer å fange opp effekten fra perioderesidualene. Vi antar modellen:

$$\log\left(\frac{q_x}{1 - q_x}\right) = \theta_{0x} + \theta_{1x}e^{\theta_{2x} \cdot x}, \quad (9)$$

for periodedødeligheten. For å kunne trekke ut perioderesidualene, må vi ha at residualene må følge:

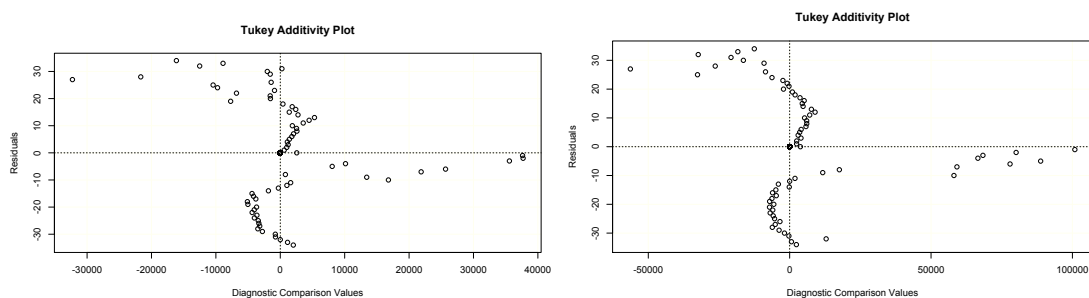
$$\epsilon_x = \log\left(\frac{\hat{q}_x}{1 - \hat{q}_x}\right) - \hat{\theta}_{0x} - \hat{\theta}_{1x}e^{\hat{\theta}_{2x} \cdot x}, \quad (10)$$

der vi har at $\hat{\theta}_{0x}$, $\hat{\theta}_{1x}$ og $\hat{\theta}_{2x}$ er de estimerte Gompertz-Makeham parametere for periodedødeligheten, \hat{q}_x er de estimerte dødelighetene og x er alder. Tilsvarende kan vi også regne ut kohortresidualene, ved simpelthen innsette

$$\underbrace{\tilde{\epsilon}_{xk}}_{\text{kohortresidual}} = \epsilon_{xk-x},$$

inn i ligning 10.

For å se etter kohorteffekter i residualene for periodedødeligheten kan vi først bruke en innebygd metode i R, nemlig `medpolish`.⁵ Denne baserer seg på ”Tukey’s Median polish”. Metoden fjerner additiv effekt av alder (rad) og år (kolonne) iterativt ved å trekke median verdi av hver rad og kolonne. Metoden fortsetter å iterere helt til man enten treffer maks antall iterasjoner (bestemt), eller kommer frem til en proposjonal reduksjon i summen av de absolutte residualer som er mindre enn en fastsatt verdi (eksempel: $\text{eps} \leq 0.1$). Fordelen med denne metoden er at resultatene kan oppnås uten bruk av matematisk transformasjon, eller transformasjon av risiko og rater, men det bør bemerkes at logaritmisk transformasjon av risiko og rater før kjøring av selve metoden kan resultere i en interaksjon på den multiplikative skala (eller logaritmisk additiv effekt), og man bør sjekke dette. Metoden inneholder også flere kombinasjoner av plott, og under er det vist plott av residualene gjennom metoden:



Figur 4.8: Kohort effekt i residualer gjennom `medpolish`, menn til venstre og kvinner til høyre

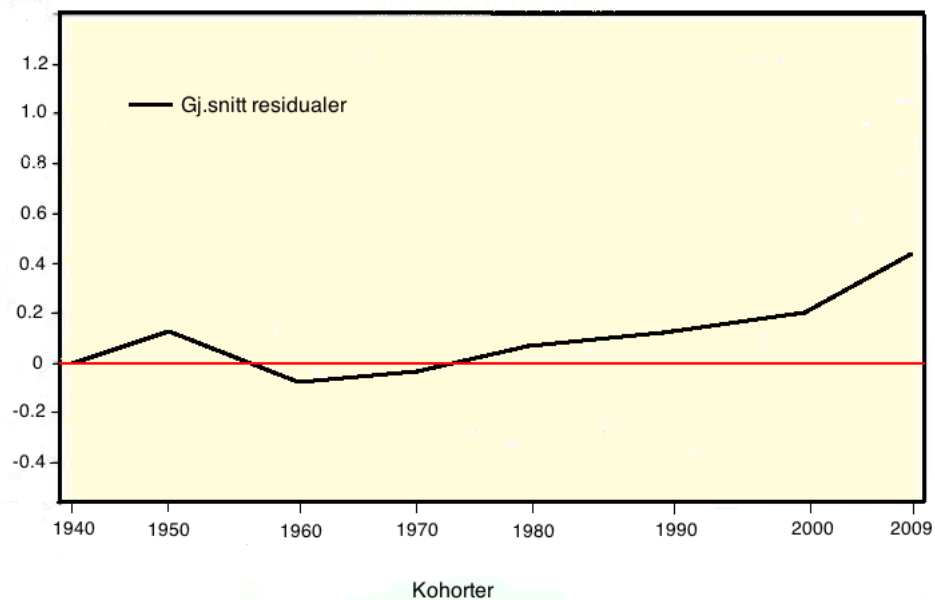
Metoden tar inn en matrise av rader (alder) og kolonner (år) med dødelighetsrater for tilsvarende plass, og gir ut median residualer. Det er altså ikke residualer vi sender

⁵<https://stat.ethz.ch/R-manual/R-patched/library/stats/html/medpolish.html>

inn, men alder og rater opp mot kohortene. Etter flere iterasjoner stabiliseres residualverdiene (rad og kolonne-medianene appropksimerer seg mot 0) og det kan sees på at disse inneholder kohorteffekten. I figuren over ser vi residualene plottet under metoden angitt som et diagnostisk verktøy. Resultatet tolkes slik;

Hvis residualene ligger på x og y-aksen som tilsvarer 0, så er det ingen kohort-effekt.

Figur 4.8 viser residualene følge x og y-aksen rundt 0 for både kvinner og menn, men vi legger merke til at vi har estimerte punkter som ligger litt lengre unna aksene. Betyr dette at vi faktisk *har* kohorteffekt? Det kommer helt ann på størrelsen av effektene. Vi ser at noen residualer differer fra aksene, men om denne differansen er stor eller liten gjenstår å se.



Figur 4.9: Residual gjennomsnittsverdier fra *medpolish* metoden, *menn*

Plott av residualene mot kohortene er en effektiv fremgangsmåte for å teste tilstedeværelse og størrelse av kohorteffekter. Hvis det ikke er noen kohorteffekt, så har residualene en tendens til å fordele seg rundt 0, fordi vi har at

$$E[\epsilon_x] = 0,$$

forventningen skal være tilnærmet 0. Avvik fra denne antagelsen kan indikere at det er kohorteffekter til grunn. Fra Figur 4.9 får vi faktisk avvik, tilsvarende avvik er også

for kvinner. Dette legger igjen til grunn for at vi har kohorteffekt i dataene for periodedødelighet. For null kohorteffekt ville vi forventet en horisontal linje på akse $y=0$, men siden den både går opp og ned rundt 0, betyr det at vi har positive og negative kohorteffekter. Størrelsen på denne effekten ser ikke ut til å være stor, og de største effektene ser ut til å gå fra periode 1975 og oppover. Dette kan igjen være et resultat av at vi har lite data for kohorter over 1980 (som forklart tidligere), som gir en større kohorteffekt i periodene og som igjen sterkt kan påvirke resultatene vi får. Vi sender inn logaritmiske dødelighetssannsynligheter inn i metoden, noe som tilsier at vi har en logaritmisk additiv effekt (og ikke på den multiplikative skalaen som forklart tidligere kunne skje).

Hittil har vi bare testet for kohorteffekt gjennom metoden `medpolish`. Ved ligning 10 modellerer vi residualer for periodedødeligheten, for hver alder 18-100, for hvert eneste år.

```

1 Gir inn:  $\hat{q}_x$ ,  $\hat{\theta}_0$ ,  $\hat{\theta}_1$ ,  $\hat{\theta}_2$ 
2
3 #tomme matriser for menn og kvinner
4  $\epsilon_{xm} = \text{matrix}(\text{NA}, \text{nrow}=\text{antall rader (alder)}, \text{ncol}=\text{antall kolonner (år)})$ 
5  $\epsilon_{xk} = \text{matrix}(\text{NA}, \text{nrow}=\text{antall rader (alder)}, \text{ncol}=\text{antall kolonner (år)})$ 
6
7 #iterer gjennom rader for menn og kvinner fra datasett
8
9 for ( i in 1: (nrow)) {
10   for (j in 1: (ncol)) {
11      $\epsilon_{xm}$           #ligning 10
12      $\epsilon_{xk}$           #ligning 10
13   }
14 }
15 Returnerer  $\epsilon_{xm}$ ,  $\epsilon_{xk}$ 

```

Algoritme 4.1: Pseudokode residual-estimering

Ved denne algoritmen modellerer vi residualene for periodedødelighetene i ϵ_x , og modellerer videre igjen kohort residualene i $\tilde{\epsilon}_{xk}$, fra deres respektive datasett. Da har vi flere metoder om hvordan vi kan gå frem for å se etter effekter i residualene. Først og fremst kan vi plote kohortresidualene mot periodeåret. Vi har fra tidligere at:

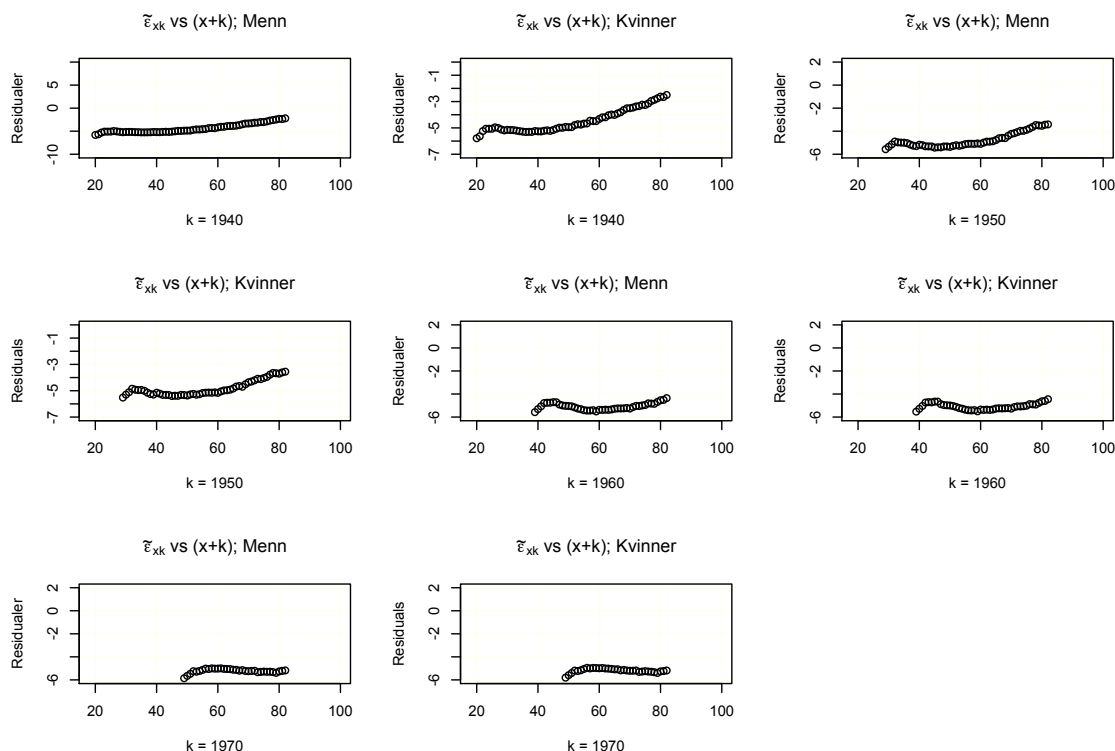
$$\underbrace{\tilde{\epsilon}_{xk} = \epsilon_{xk-x}}_{\text{kohortresidual}}$$

Ved å plote $\tilde{\epsilon}_{xk}$ mot periodeåret vil si at vi matematisk plotter $\tilde{\epsilon}_{xk}$ mot $x+k$ (alder + kohortår). Men dette førte til at iterasjon av dataene gikk utenfor dets grenser ($x+k > \text{antall rad/kolonne}$). Så enda en restriksjon måtte til:

Restriksjon 2: $(x+k) \leq \text{siste rad i residual-datasett for hver kolonne}$.

4.3 Minste kvadraters metode og residualer

Restriksjonen fører til at de eldste eldre uteblir, men får grafisk fremstilt kohortresidual-ualet mot periodeåret, og se om vi finner noe mønster eller effekter i dette. I tillegg angir vi k ,



Figur 4.10: Utvalg av $\tilde{\epsilon}_{xk}$ vs. $(x+k)$, for menn og kvinner

slik at denne er fast og simulerer det simultant for alle x (aldre). Vi ser et lite mønster i figuren ovenfor, men vi klarer ikke fastsette størrelsen eller effekten ut av mønsteret fra kun dette. Vi må enda dypere inn i analysen av residualene.

4.3 Minste kvadraters metode og residualer

Minste kvadraters metode er en standard metode for omtrentlige løsninger av overdeterminerte systemer, det vil si; system av ligninger der det er flere eller like mange ligninger som ukjente variable. "Minste kvadrat" betyr at den totale løsningen minimerer summen av kvadratene av de feilene som er gjort i resultatene fra hver enkelt ligning. Den viktigste applikasjonen er innen datatilpasning. Den beste tilpasning i minste kvadraters forstand minimerer summen av kvadrat-residualene, og det er dette vi skal teste med våre residualer.

Vi har at residualene $\tilde{\epsilon}_{xk}$ (for kohorter) gis gjennom ligning 10. Ved å sette opp en minste kvadrates metode for residualene våre, kan vi bestemme a og b slik at:

$$\sum_{k=k_0}^K \sum_{x=x_0}^X n_{xk} (\tilde{\epsilon}_{xk} - a - b(x+k))^2, \quad (11)$$

der vi har at n_{xk} er lik antall individer utsatt for risiko for å dø (som tidligere). Dette gir en løsning av:

$$(i): \mathbf{a} \left(\sum_k \sum_x n_{xk} \right) + \mathbf{b} \left(\sum_k \sum_x (x+k) \right) = \sum_k \sum_x n_{xk} \cdot \tilde{\epsilon}_{xk},$$

$$(ii): \mathbf{a} \left(\sum_k \sum_x n_{xk} \cdot (x+k) \right) + \mathbf{b} \left(\sum_k \sum_x n_{xk} \cdot (x+k)^2 \right) = \left(\sum_k \sum_x n_{xk} \cdot (x+k) \right) \cdot \tilde{\epsilon}_{xk},$$

og ved å tilegne variable til ligningene over har vi:

$$\begin{aligned} A_{11} &= \sum_k \sum_x n_{xk}, \\ A_{12} &= A_{21} = \sum_k \sum_x n_{xk}(x+k), \\ A_{22} &= \sum_k \sum_x n_{xk}(x+k)^2, \\ B_1 &= \sum_k \sum_x n_{xk} \cdot \tilde{\epsilon}_{xk}, \\ B_2 &= \sum_k \sum_x n_{xk} \cdot (x+k), \end{aligned}$$

slik at vi får det enkle systemet av 2 ligninger med 2 ukjente:

$$\begin{aligned} A_{11} \cdot \mathbf{a} + A_{12} \cdot \mathbf{b} &= B_1, \\ A_{21} \cdot \mathbf{a} + A_{22} \cdot \mathbf{b} &= B_2. \end{aligned}$$

Gjennom dette oppsettet kan vi enkelt bruke innebygde systemer i R til å løse a og b i ligningene med hensyn på datasettet vårt, separat for både kvinner og menn.

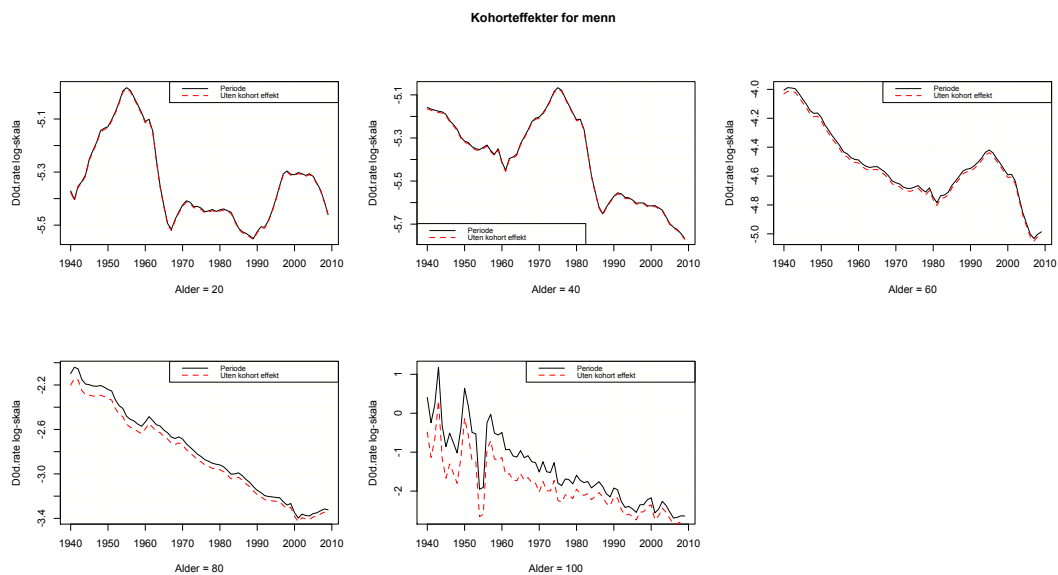
4.3 Minste kvadraters metode og residualer

	Menn	Kvinner
a	-1.490169e-05	-9.168559e-05
b	2.056005e-07	1.204356e-06

Tabell 4.2: Beregninger for variable a og b i minste kvadraters metode, for menn og kvinner

Før vi oppnådde resultatene var det antatt at a ville ligge nærme 0, noe som stemmer. a -verdien er negativ for begge parter. Det som er interessant å se er at b -verdien for både menn og kvinner også er veldig liten. Betyr dette at størrelsen på kohorteffekten er lav? For å determinere hva verdiene av a og b faktisk betyr, må vi foreta enda en test. Ved å omformulere ligning 9, kan vi plote dette mot periodeåret, og se hva størrelsen på variablene innebærer, på log-skalaen.

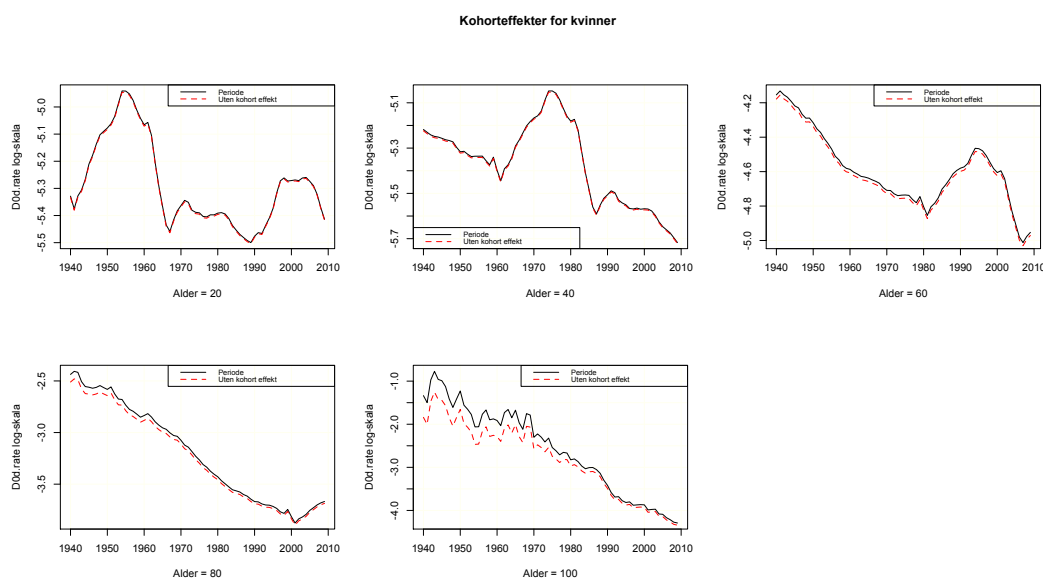
$$\log\left(\frac{q_x}{1-q_x}\right) = \theta_{0x} + \theta_{1x}e^{\theta_{2x} \cdot x} \implies \log\left(\frac{\tilde{q}_{xk}}{1-\tilde{q}_{xk}}\right) = \theta_{0k} + \theta_{1k}e^{\theta_{2k} \cdot x} + \underbrace{(a + b(x+k))}_{\text{angir størrelsen}}.$$



Figur 4.11: Utvalg av observerte dødelighetsrater for visse aldre mot baner med og uten kohorteffekt, på log-skala, *menn*

Endelig har vi kommet frem til en metode for å se størrelsen på selve kohorteffekten. Her har vi dødelighetsrater på y-aksen (log-skala) mot periodeår på x-aksen, for en gitt alder. Den røde stiplede linjen angir raten for periodeåret uten selve kohorteffekten. Vi

noterer oss med en gang at generelt er dødelighetene til individer med gitt alder lavere uten en kohorteffekt. For alder lik 20 og 40 ligger kurvene så og si rett på hverandre, men vi begynner å se en forskjell for alder 60 og oppover (minner om at figurene her er forminsket for å passe formatet, og at forskjellen er større ved normal størrelse på figur). Det er tidligere forklart at vi ikke har fullstendige kohorter for de eldste aldre, da alle individer i disse kohortene ennå ikke har dødd ut. Det er da rimelig at den stiplede linjen ligger lavere enn den svarte, fordi den svarte bygger kun på dødeligheten til individer med lik alder fra siste år, mens kohortene trenger dødeligheten til alle individene. Effekten er ikke altfor stor, men det *er* tross alt en effekt her. Tilsvarende effekt får vi også for kvinner.



Figur 4.12: Utvalg av observerte dødelighetsrater for visse aldre mot baner med og uten kohorteffekt, på log-skala, *kvinner*

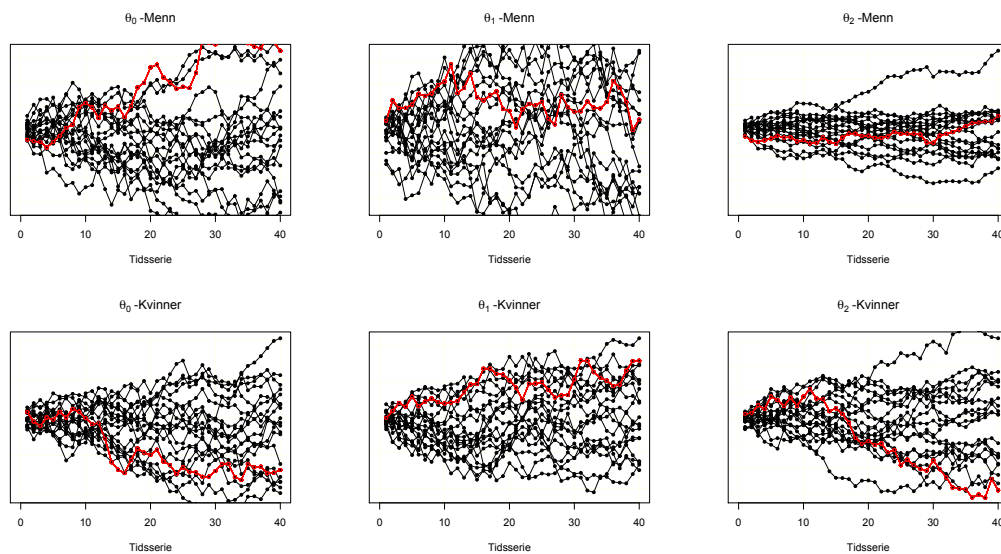
4.4 Dynamisk modell og simulert utvikling

Vi vet nå at Gompertz-Makeham modellen klarer å fange opp kohorteffekten i periodedødeligheter. Gjennom *ligning 7* har vi modell for dødelighetene, og i *avsnitt 4.1* ble det vist at det kan estimeres gode parametere i Gompertz-Makeham modellen, for både kohort- og periodedødelighetene. Før vi prøver oss på en økonomisk anvendelse, må vi se hvordan parametere vil se ut med hensyn på fremtidsmessig progresjon. Vi vil altså frem til en dynamisk modell som kan håndtere livsforsikringsberegninger for kohorter. Det som skal gjøres videre er da å bruke beregningen av parameterverdiene gitt ved figur 4.1 og 4.2 til å simulere utviklingen av disse parameterverdiene. Vi tester for korrelasjon i parameterene:

Menn			Korrelasjon		
	Gj.snitt	sd.avvik	θ_0	θ_1	θ_2
θ_0	0.0048	0.0001	1	-0.2565	0.6247
θ_1	5.22 e-05	2.67 e-05	-0.2565	1	-0.8874
θ_2	0.1160	0.0036	0.6247	-0.8874	1
Kvinner			Korrelasjon		
	Gj.snitt	sd.avvik	θ_0	θ_1	θ_2
θ_0	0.0049	0.0002	1	-0.8215	0.9328
θ_1	7.71 e-05	3.68 e-05	-0.8215	1	-0.6686
θ_2	0.1035	0.0046	0.9328	-0.6686	1

Tabell 4.3 : Gjennomsnitt, standardavvik og korrelasjon mellom parametrene θ_0 , θ_1 og θ_2 i Gompertz-Makeham modellen, for menn og kvinner, *kohortdata*

Fra tabell 4.3 ser vi at det er ganske høy korrelasjonsfaktor mellom parameterene, for både kvinner og menn. Menn har stor korrelasjon mellom θ_0 og θ_1 til θ_2 , mens for kvinner er det enda større korrelasjoner mellom parameterene. Vi kan nå konstatere med at det er avhengighet mellom parameterene. Vi kan tenke oss at utviklingen av parameterene i Gompertz-Makehams modell er av en slags form for en tidsrekke. Med dette i fokus vil vi da sette opp en stokastisk modell for θ -ene.



Figur 4.13: Random-walk test for Gompertz-Makeham parameterene for menn og kvinner i tabeller Appendix B2 og B3

En stokastisk prosess er en kolleksjon av tilfeldige variable knyttet til tidspunkter. Hvis variablene ikke har noen naturlige grenser (kan fluktuere fritt), så er disse svingningene av en natur som karakteriserer en random-walk/tilfeldig-gange. Ved å estimere standardfeilen av differansen mellom parameterene, med forventning lik 0, kan vi anslå en test for sjekk av random-walk. I figur 4.13 har vi 20 realisasjoner av denne typen opp mot parameterene, og vi ser at θ -ene ikke avviker fra realisasjonene, slik at vi kan anta at vi har en prosess av random-walk. Den simpleste formen av dynamikken i en random-walk er:

$$X_t = X_{t-1} + \xi_t,$$

hvor $\xi_t \sim \text{IID}(0, \sigma^2)$ er støy-ledd gitt som standardnormalfordelte variable. Vi får i vårt tilfelle at:

$$\begin{aligned}\theta_{0,k} &= \theta_{0,k-1} + \xi_{0,k}, \\ \theta_{1,k} &= \theta_{1,k-1} + \xi_{1,k}, \\ \theta_{2,k} &= \theta_{2,k-1} + \xi_{2,k},\end{aligned}\tag{12}$$

hvor k er kohort.

Q-Q test av støy-ledd fra historiske data fastslår at vi får noen tyngre haler for noen av støy-leddene, både for menn og kvinner. Dette kan komme av korrelasjonen vi får mellom parameterene våres. Fordi det i historikken vises korrelasjon mellom θ -ene vil vi også ha med dette under simuleringen av prediksjonene for årene som kommer, og således blir dette inkludert under simuleringen av $\xi_{i,k}$ -ene (støy-leddene). På denne måten beholder vi avhengigheten mellom parameterene under fremskrivningen. Lineært avhengig støy-ledd trekkes ved følgende metode.

Støy-leddet ξ er som sagt standardnormalfordelt. Da har vi at den ubetingede sannsynlighetstetthetsfunksjonen følger

$$f_{\xi_i}(x) = \frac{1}{\sqrt{2\pi k}} \cdot e^{-\frac{x^2}{2k}},$$

for en gitt k , og $i = 0, 1, 2$. Av dette fremkommer:

$$\begin{aligned}E[\xi_i] &= \zeta_i, \\ \text{var}(\xi_i) &= \sigma_i,\end{aligned}\tag{13}$$

og videre at kovariansen mellom ξ_i og ξ_j gis ved:

$$\text{cov}(\xi_i, \xi_j) = \sigma_{ij}.$$

Ved å bruke at

$$\text{cov}(\xi_i, \xi_i) = \text{var}(\xi_i) = \sigma_{ii} = \sigma_i,$$

kan vi sette opp kovariansmatrisen $\Sigma = \sigma_{ij}$:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}.$$

Denne matrisen kan løses ved hjelp av Choleskydekomposisjon. Vi vil ha ut koeffisientene $\mathbf{C} = (c_{ij})$, og en standard metode er å sette en restriksjon $c_{ij} = 0$, når enn $j > i$.⁶ Choleskymatrisen har som egenskap at alle elementene over diagonalen vil være lik 0. Matrisen vil alltid tilfredsstill

$$\sigma_{ij} = \text{cov}(\xi_i, \xi_j) = \sum_{n=1}^N c_{in} \cdot c_{jn},$$

og følgende kriteriet:

$$c_{i1}^2 + c_{i2}^2 + c_{i3}^2 = 1, \quad \forall \quad i=1,2,3.$$

Cholesky-matrisen blir som følger:

$$\mathbf{C} = \begin{pmatrix} c_{11} & 0 & 0 \\ c_{21} & c_{22} & 0 \\ c_{31} & c_{32} & c_{33} \end{pmatrix}.$$

Vi har at støyledd-variablene var standardnormal, uavhengig og identisk fordelte. Cholesky-matrisen brukes til å konvertere variablene til å bli korrelerte, slik vi ville ha dem fordi parameterene våre historisk også var korrelerte. For at det skal skje så trekker vi standardnormalfordelte variable som er uavhengig og identisk fordelte

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix},$$

og setter $\boldsymbol{\psi} = (\psi_{1,k}, \psi_{2,k}, \psi_{3,k})$, slik at:

$$\boldsymbol{\psi} = \mathbf{C} \bullet \boldsymbol{\eta}, \tag{14}$$

hvor \mathbf{C} er Cholesky-matrisen. Tilslutt får vi endelig

⁶side 149, Computation and Modelling in Insurance and Finance

$$\begin{aligned}\xi_{0,k} &= \zeta_0 + \psi_{0,k}, \\ \xi_{1,k} &= \zeta_1 + \psi_{1,k}, \\ \xi_{2,k} &= \zeta_2 + \psi_{1,k},\end{aligned}\tag{15}$$

hvor ζ er gitt ved ligning 13. Ved å bruke ligning 14 og 15 kan vi endelig simulere korrelerte normalfordelte ξ , som da vil tilfredsstille korrelasjonen gitt mellom parameterene i tabell 4.3.

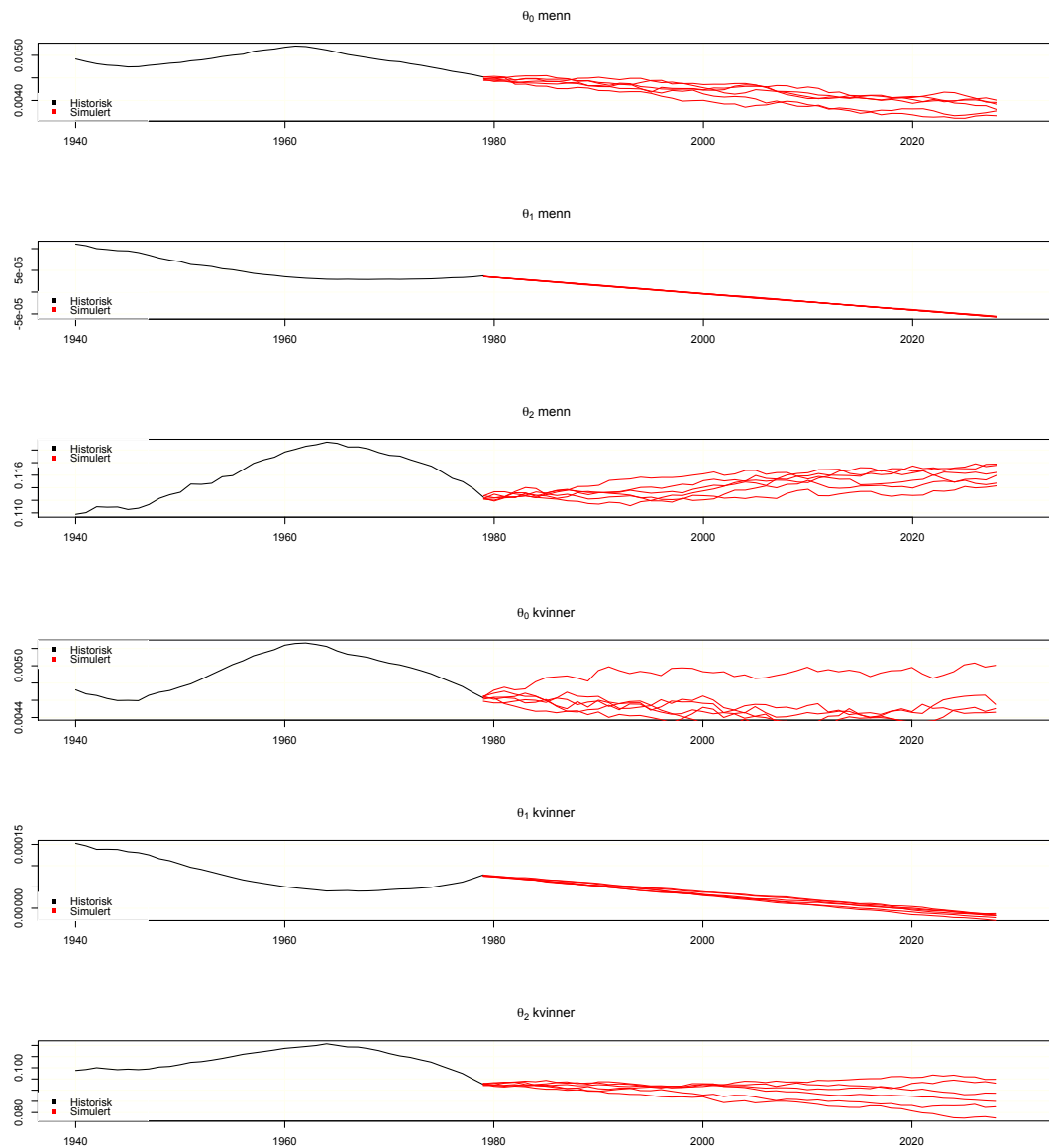
```

1  Gir inn: C,  ζ0, ζ1, ζ2, n                                #simulerer "n" antall ganger
2
3  for( k in 1:50){                                           #simulerer 50 år frem
4      for(i in 0:3){
5          trekker η ~ N(0,1)                                #uavhengig normalfordelte
6      }
7      for(j in 0:3){
8          ψj = 0
9          for(i in 0:3){
10             ψj = ψj + cj+1,i+1 · η                      #ligning 14
11         }
12         ξj,k = ζj + ψj                                    #ligning 15
13     }
14     Returnerer ξ0,k, ξ1,k, ξ2,k
15
16 }
```

Algoritme 4.2: Pseudokode for å simulere utvikling av parametere

Figur 4.14 viser da utviklingen i Gompertz-Makeham parameterene våres. Vi har i dette tilfellet simulert 50 år frem, og vi ser en del variasjon i utviklingen. Spredningen blir større og større jo lenger frem i tid vi simulerer, fordi det er umulig å spå utviklingen i fremtiden eksakt.

4.4 Dynamisk modell og simulert utvikling



Figur 4.14: Utvalg av simulerte baner for utviklingen av GM-parametere

Oppsumert

Vi har til nå tilpasset en dynamisk modell som passer kohortdataene våres. Vi har modellert parametere i Gompertz-Makehams modell, som vi så passer godt mot rådataene, både i kohort- og periodetilfellet. Det viktigste resultatet til nå har vært beregningen av kohorteffekten. I figur 4.11 og 4.12 så vi selve størrelsen på kohorteffekten opp mot periodedødeligheten. For kohorter lengre tilbake i tid var denne effekten minimal, men for nyere kohorter var denne større. Vi forklarte tidligere at dette kommer av mengden av data vi har tilgjengelig. Eldre kohorter har mer data for aldre opp mot 100, mens nyere kohorter er langt ifra fullstendige, da individene i disse kohortene ennå er unge. Dette er et resultat vi tar med videre, og som vil brukes for å se hva slags økonomiske konsekvenser dette gjenspeiler.

5 Pensjonsberegninger

I praksis er det flere kilder til usikkerhet i livs- og pensjonsforsikringer. I seksjon 1.4 ble det gitt eksempler på at menn og kvinner i vestlige land lever lenger og lenger. Dette, og at det har blitt en nedgang i dødelighet, fører til systematiske feil i modelleringen, fordi dødeligheten til individer som er forsikret eller er en del av et pensjonsprogram ofte er lavere enn gjennomsnittet, noe som igjen fører til underestimering av forpliktelsene til forsikrings og pensjons-giveren. Denne gruppen av individer er en del av et ugunstig utvalg (*adverse selection* på engelsk), og gir selskaper store problemer når det kommer til riktig modellvalg og estimering av premier, fordi det er en relativ stor kilde til usikkerhet.

Variable vi definerer skal være kjønnsnøytrale, altså like for menn og kvinner. Menn og kvinner har ulike dødeligheter som kan lede til ulike forsikringer, men fordi Likestillingsombudet bestemte at det i lovverket skal være kjønnsnøytralt mellom kvinner og menn, så må selskaper tilrettelegge seg etter dette. Det nevnes at var det en stor sak mellom Likestillingsombudet og forsikringsbransjen for et par år siden, der det ble krevd at det ble gitt like forsikringer til unge førere for begge kjønn, selv om det var større risiko for unge menn å skade bilen enn kvinner da unge menn generelt er mer fartsglade. Dette førte til at forsikringsselskapene økte forsikringspremiene til kvinner på lik linje med menn, selv om alle beregninger og estimater pekte mot lavere forsikring for unge kvinner.

5.1 Innledning og antakelser

Kvantitet	Symbol	Størrelse
Rente	r	$r = 3\%$
Start alder	l_0	$l_0 = 18$
Maks alder	l_e	$l_e = 100$
Pensjonsalder	l_r	$l_r = 67$
Diskonteringsfaktor	d	$d = \frac{1}{1+r}$
Utbetaling	s	$s = 1$
Premie	π	

Tabell 5.1 : *Variable i våre beregninger*

Vi har nå satt opp en dynamisk modell for kohortdataene, og vi har sett at modeller av denne typen gir mye bedre beregninger og anslag enn en statisk modell med faste parametere. De økonomiske anvendelsene er noe av det grunnleggende et forsikrings-selskap er interessert i, og vi skal teste hva slags økonomiske konsekvenser vår modell kommer frem til. Et forsikringsselskap har forpliktelser ovenfor sine kunder, og vi skal se om kontanstrømmen inn og ut av pengekassen/pensjonskassen er annerledes ved bruk av kohortdata, og i tillegg se om kohorteffekten beskrevet i seksjon 4.3 spiller en rolle

på selve kontanstrømmen. Vil nedgang i dødelighet for nyere kohorter gi annerledes resultater enn dødeligheten for de eldre kohortene? Variabler vi bruker er beskrevet i tabell 5.1. Ved å sette $s = 1$, altså en konstant som skal beskrive utbetalingene, kan vi heller få nivået for utbetalinger enn å få beløpet i kroner.

Vi skal se på verdien av poliser. Da må vi ha en diskonteringsfaktor $d = \frac{1}{1+r}$ som gir nåverdien av en fremtidig pengesum. Dette er fordi penger mister verdi over tid. Denne faktoren tar inn renten i nevneren fra tabell 5.1, og vi ser bort ifra inflasjon. Anta en pengestrøm som stopper når et individ dør. I pensjonsskjema (livsforssikringer) er det ofte pengebidrag $\{\pi\}$ (negative) opp til pensjonsalderen (inntruffet død), og utbetalinger $\{s\}$ (positive) etter inntruffet pensjonsalder (død). Vi antar at det kun er to mulige tilstander for individet: Levende eller død (vi utelukker uførhet og andre muligheter). Vi beskriver dynamikken for premier på polisenivå (altså for et enkelt individ).

Tilstand 1: Aktiv/Pensjonert

Tilstand 2: Død

En person i alder l vil da få følgende oppsett:

$$\begin{array}{ll} s_l(1) = -\pi & l < l_r \quad \text{betaler premie,} \\ s_l(1) = s & l > l_r \quad \text{mottar utbetaling,} \\ s_l(2) = 0 & \text{personen er død.} \end{array}$$

Netto engangspremie er størrelsen på en pengesum som må hentes når en polise blir utstedt ved en alder l_0 for å sikre at det skal være nok penger til å betale pensjonsutbetalingen (livsforsikringen) av polisen. Nåverdien er gitt ved formen:

$$PV_0 : \sum_{k=0}^K = d^k \cdot \lambda_k, \quad (16)$$

hvor λ_k er en sekvens av betalinger, k er kohort, d er diskonteringsfaktoren og PV_0 står for nåverdien. Alle betalinger inn og ut av polisen er balanserte, slik at dets forventede verdi er lik null. For forsikreren er det ingen profitt å tjene fra dette, ergo kommer selskapets inntekt gjennom rentesatsen r . Ekvivalensprinsippet justerer premium (forskuddsbetalinger) slik at det passer utbetalingen gitt ved senere tid. Hvis π er premie og s er utbetaling, så har vi simpelt at:

$$\underbrace{E[\lambda_k] = -\pi \cdot {}_k p_l}_{\text{premiebetalinger}} \quad \text{og} \quad \underbrace{E[\lambda_k] = s \cdot {}_k p_l}_{\text{utbetalinger}}$$

hvor ${}_k p_l$ er overlevelsessannsynligheter fra ligning 1 i avsnitt 2.2. Bryter vi ligning 16 ned til vårt tilfellet så får vi:

$$E[PV_0] = -\pi \sum_{k=0}^{l_r-l_0-1} d^k \cdot {}_k p_{l_0} + s \sum_{k=l_r-l_0}^{100} d^k \cdot {}_k p_{l_0}, \quad (17)$$

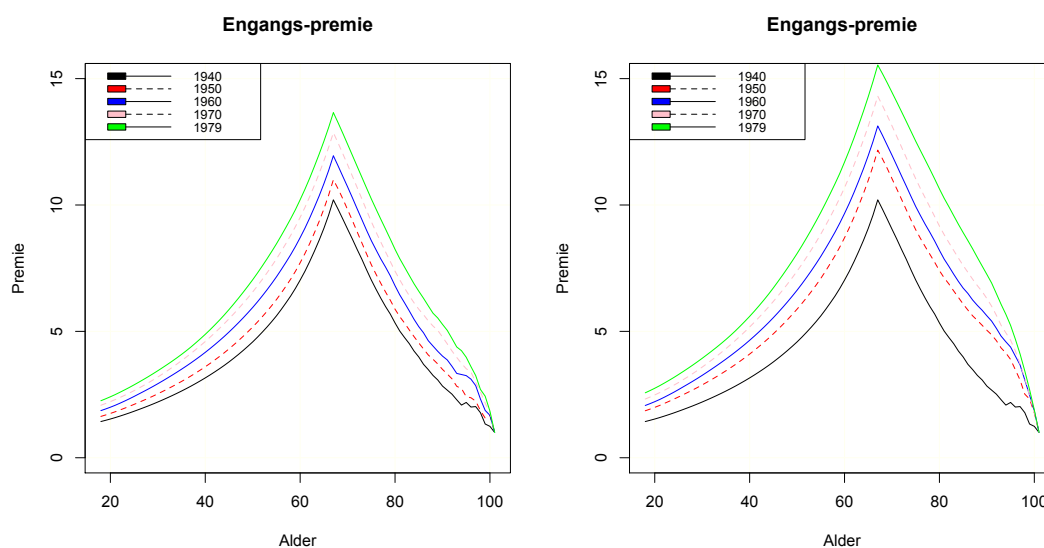
hvor vi har l_0 , l_r , s og d fra tabell 5.1. For å tilfredsstille ekvivalensprinsippet har vi at $E[PV_0] = 0$, og slik blir premien π også kalkulert.

5.2 Beregningsmetode

For å kalkulere π må vi videre gjennom 2 steg. Først må vi beregne ${}_k p_l$ fra dødelighets sannsynlighetene, gjennom ligning 1. I statistisk terminologi prøver vi altså å konstruere en livstabell av overlevelsessannsynligheter fra kohorter. Dette gjøres ved å foreta rekursjonen (omgjør ligning 2):

$${}_{k+1}p_l = \{1 - q_{l+k}(l)\} \cdot {}_k p_l, \quad \text{med startpunkt } {}_0 p_k = 1.$$

Når livstabellen er konstruert er det kun å modifisere en funksjon som kan ta inn denne matrisen, og som beregner π for oss gjennom ligning 16.



Figur 5.1: Utvalg av premie for kohorter med pensjonsalder 67, menn til venstre og kvinner til høyre

For hver eneste alder 18-100 år er det simuleringer som går over overlevessannsynlighetene og anslår premien. Vi har satt pensjonsalderen til 67 år (denne kan selvfølgelig endres). Fordi kohortene våre ikke er fullstendige for alle aldre, så er det simulert fremskrivinger av premium for de aldrene som mangler i hver kohort.

I figuren ovenfor ser vi hvordan premien varierer for de forskjellige kohortene. De yngre kohortene har dødeligheter som er lavere enn de eldre kohortene og må betale høyere premie. Dette kommer som forklart av at individer lever lenger og lenger. Med nedgang i dødelighet så er en økning i premie nødvendig for at **ligning 17** skal balanseres. *Det koster å leve lenger.* I kapittel 3 så vi at kvinner generelt har lavere dødelighet enn menn på alle alderstrinn, som igjen fører til at de må betale høyere premie enn menn, og som også fremkommer av figuren.

Vi har altså estimert single premier for både menn og kvinner fra kohortene, og resultatene ser gode ut med tanke på at de skal ta inn det faktum at forskjellige kohorter har forskjellige dødeligheter. Det som gjenstår å se er hvordan disse resultatene endres om vi inkluderer (utelater) kohorteffekten vi kom frem til i avsnitt 4.3. Vil premiene endre seg stort eller minimalt om vi utelater denne effekten?

```

1  Gir inn:  $\tilde{q}_{xk}$ ,  $l_e$ ,  $l_r$ ,  $l_0$ ,  $s$ ,  $d$ 
2
3   ${}_k p_l$                                 #tom matrise
4   ${}_o p_l = 1$                             #initial betingelse
5
6  for (l in 18: 100){                      #rekursjon
7    for (t in 2: antall år){
8       ${}_{t+1} p_l = (1 - \tilde{q}_{l+t}(l)) * {}_t p_l$     #beregner livstabell
9       ${}_{t+1} p_l = (1 - \tilde{q}_{l+t}(l)) * {}_t p_l - (a + b(1 + t))$  #uten kohorteffekt
10   }
11 }
12 enganspremie=function(s, l_r, d){
13
14    $l_e = \text{length}(p_l)$ 
15    $kp = {}_k p_l$ 
16    $I = \text{matrix}(0, l_e + 1, l_e + 1)$           #indikator matrise
17    $I[\text{row}(I) + \text{col}(I) > l_r + 1] = 1$ 
18
19    $ll = 0:l_e$                                 #beregner premium
20    $M = s * (d ** ll) * kp * I$ 
21    $pi = \text{apply}(M, 2, \text{sum})$ 
22    $\text{list}(pi = pi)$ 
23 }
24 Gir ut:  $\pi$ ,  $\bar{\pi}$ 

```

Algoritme 5.1: Pseudokode for beregning av premium med og uten kohorteffekt

Vi modellerte minste kvadraters metode på residualene for kohorter i avsnitt 4.3, for så å plote dette ut mot periodeåret som viste oss kohorteffekten gjennom tabell 4.2. En metode å gå frem på er å implentere parameterene a og b slik at de tas inn sammen med kohortdødelighetene når livstabellene blir konstruert. Det er kun hvis vi setter inn effekten i dødelighetene at vi kan se etter en eventuell forskjell i premiene, og vi

5.3 Resultater

må derfor ha at denne effekten må gå gjennom hele prosessen. Fra kohortdødeligheten blir livstabellene med parameterene konstruert, og fra livstabellen vil igjen premiene bli estimert. Vi har fra tidligere at kohortdødeligheten gis ved \tilde{q}_{xk} .

Modellering gjennom algoritme 5.1 gir oss π og $\bar{\pi}$, henholdsvis premier med og uten kohorteffekt, respektivt. Tabell 5.2 viser til en oversikt over premienivået for utvalgte kohorter, for alder $\in \{30, 60 \text{ og } 90\}$. Premieverdien henviser ikke til kroner, da vi innledningsvis valgte å sette konstanten s lik 1, slik at vi får nivået for premier istedet.

5.3 Resultater

Premie						
	1940			1950		
	30	60	90	30	60	90
Menn	2.2/ 2.1	7/ 6.9	2.8/ 2.8	2.5/ 3	7.7/ 8.2	3.5/ 3.9
Kvinner	2.5/ 2.7	8/ 8.4	4.2/ 4.1	2.9/ 3	8.7/ 9.2	5/ 4.5
	1960			1970		
	30	60	90	30	60	90
Menn	2.9/ 3.9	8.7/ 9.7	4.0/ 5.0	3.2/ 4.7	9.5/ 11	4.8/ 6.2
Kvinner	3.2/ 3.8	9.6/ 10.6	5.6/ 5.4	3.6/ 4.6	10.7/ 12	6.3/ 6.6
	1979					
	30	60	90			
Menn	3.4/ 5.4	10.2/ 12.2	5.3/ 7.2			
Kvinner	3.9/ 5.5	11.7/ 13.4	7.2/ 8.3			

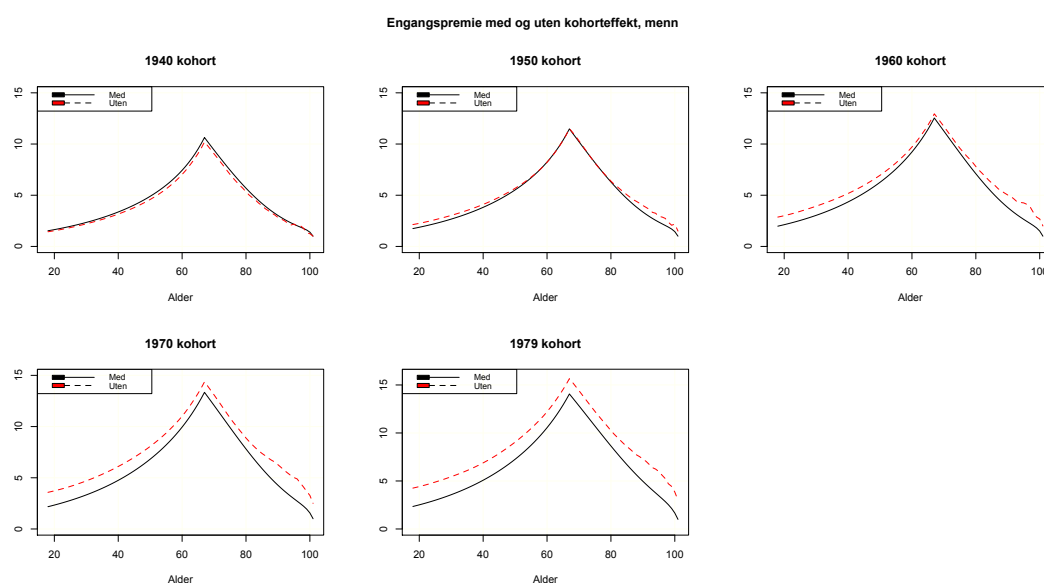
Tabell 5.2: Premie-beregninger for utvalgte kohorter og aldre, med og uten kohorteffekt. Uten kohorteffekt i **rødt**

Fra figur 4.11 og 4.12 så vi at dødelighetene var lavere for kohortene uten kohorteffekt. Dette ble også mer synlig jo yngre kohortene var. Analyse av figur 5.1 tilsa at nedgang i dødelighet impliserer høyere premiebetaling. Ergo kan vi anta at premier estimert over dødeligheter uten kohorteffekt vil gi enda høyere premier, og tabell 5.2 styrker vår antagelse.

Foruten 1940-kohorten for menn så har vi høyere premier for hver alder i alle kohorter for både menn og kvinner. Differansen på premier med og uten kohorteffekt er ikke av stor grad for de eldste kohortene, men vi ser en klar forskjell på premier for de yngre kohortene. Det er i gjennomsnitt en differanse på 2 nivåer for de yngre kohortene, en differanse som kan være av stor betydning når det kommer til forsikrings- og pensjonsberegninger. Dette styrker igjen våre resultater fra tidligere kapitler om at kohortberegninger under kohort-baserte modeller *er* annerledes enn estimeringer av

modeller basert på periodedødeligheter, og kan ha påvirkning på tilfeller der kohortinformasjonen neglisjeres. Som vi ser av tabellen ovenfor så vil neglesjering av kohorteffekten føre til lavere premiebetalingar inn til selskapene. Aktuarer og forsikringstakere verden rundt jobber med å finne optimale modeller som tar hensyn til at mennesker lever lenger og lenger. Ved å ikke ta hensyn til at kohorteffekter eksisterer kan selskaper miste potensielle økte inntekter som igjen brukes til å balansere ligning 17.

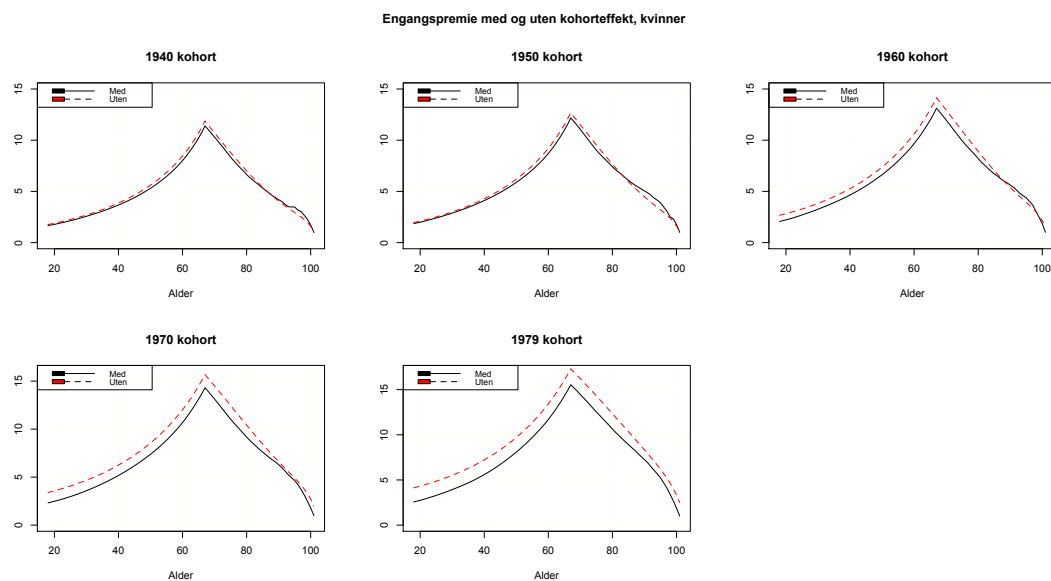
Men det hele kommer igjen tilbake til datasettet. Det var klareste differanse for de yngre kohortene, og disse er som sagt langt ifra fullstendige datasett. Ser vi på kohort 1979 så eksisterer det ikke 60 og 90-åringer idag, det er kun estimerte projeksjoner. Det var en lav forskjell i premiene for de eldre kohortene, og disse har langt flere



Figur 5.2: Utvalg av kohorter for engangs-premie med og uten kohorteffekt, *menn*

data enn de yngre kohortene. Det er også en mulighet at et komplett datasett ikke gir noen form for kohorteffekt, og at denne kun dannes når fremskrivninger skal til. I figurer 5.2 og 5.3 har vi grafisk fremstilt estimatene i tabell 5.2.

5.3 Resultater



Figur 5.3: Utvalg av kohorter for engangs-premie med og uten kohorteffekt, *kvinner*

Vi ser klart at forskjellene i premier danner seg for de yngre kohortene. Vi har i vårt tilfelle kun regnet ut nåverdien PV_0 for premier på polisenivå. Rentesatsen r ble satt fast på 3%. Denne kan modifiseres til å ta inn inflasjon, slik at den vil fluktuere med hensyn på tiden. Det nevnes at rensesatsen kan gi stor variasjon i nåverdien, med tanke på størrelse og om den er et fast inkrement eller ikke. Generelt har vi at lavere r vil gi høyere nåverdier, og motsatt. Prosedyren kan utvides, for eksempel til å se på porteføljer bestående av kohorter, men dette blir ikke gjort her.

6 Konklusjon

Vi skulle i denne oppgaven komme frem til en dynamisk modell som kunne modelleres og kjøres over kohortdødeligheter for land i Skandinavia. Fordi kohortdødeligheter oppfører seg annerledes enn periodedødeligheter måtte en slik modell tilpasses til dataene. Det ble tidlig trukket frem at Gompertz-Makeham modellen er en av de mest populære matematiske beskrivelsene av dødeligheter, og det var derfor naturlig å nettopp teste denne modellen for vår oppgave. Når det kommer til dødelighetsmodeller i bruk i dagens forsikringsverden er det periodedødeligheter som blir brukt, men få tenker på at disse stammer fra ulike kohorter, informasjon som ofte neglisjeres, og hvilke endringer som oppstår når disse ikke blir ignorert.

Vi begynte med å se etter tilpasningen av modellen med hensyn til dataene. Innledningsvis i kapittel 4 estimerte vi parametere for Gompertz-Makehams modell, og disse ble testet opp mot rådataene. Modellen passet dataene godt, noe som førte til at vi brukte denne modellen videre i vår analyse. Det ble estimert enda bedre resultater når det ble brukt tidsinkremerte parametere i motsetning til faste parametere i en statisk modell, noe vi så fra figurer gjennom kapittel 3 og 4. Spesielt ble denne forskjellen betraktelig stor når vi estimerte forventet levetid for individer (figur 3.9 opp mot figur 4.4). Simuleringer og prediksjon av fremtidig utvikling av parameterene gav større og større variasjon jo lenger frem i tid vi kom. Fremtiden kan som sagt ikke predikeres eksakt, og litt variasjon i de historiske dataene kan gi store variasjoner i utviklingen.

Videre testet vi effekter når vi ikke neglisjerte kohortinformasjonen beskrevet ovenfor. I avsnitt 4.3 brukte vi minste kvadraters metode på residualene i modellen, for å se etter en kohorteffekt i dataene. Ved å sette resultater fra metoden opp mot periodeåret fikk vi en grafisk fremstilling av effekten som ofte blir ignorert. Figur 4.11 og 12 bekreftet at kohorteffekten ble større og større jo yngre kohortene ble, gitt opp mot eldre periodeår (år nærmere vårt inneværende årstall). Det essensielle med å vise dette var at det faktisk *var* en underliggende effekt i dataene, som kun blir trukket ut og vist frem hvis man spesifikt tenker på dette. Dagens modeller som tar inn dødeligheter viser kun overflatisk fremstilling av dødelighetene, og med mindre man bevisst går inn for å finne forskjeller og effekter i dataene, så blir kohortinformasjon veldig lett oversett.

Når det gjelder de økonomiske anvendelsene så estimerte vi kun single premier i ulike aldre for begge kjønn. Vår dynamiske modell tar hensyn til forbedret levetid for individer i de ulike aldrene. Forbedret levetid er av stor bekymring for forsikringsgivere, da underestimert av premier kan gi store tap i inntekt og verdi. De eldre kohortene har mye lavere anslåtte premiebetalingen enn de yngre kohortene, noe som kommer av at nyere kohorter har lengre forventet levetid (figur 4.4). Av figur 5.1 ser vi endringen i premier for utvalgte kohorter, og av dette kan vi anta at premier kun vil øke for kohorter i kommende generasjoner. Separat for kvinner og menn har vi generelt sett lavere dødelighet for kvinner enn menn på alle alderstrinn, noe som igjen fører til at kvinner har større premiebetalingen enn menn.

Tilslutt testet vi på forskjeller i premiebetalinger, for modeller med og uten kohort-effekt. Tabell 5.2 angir premier for utvalgte individer og kohorter, både med og uten denne effekten. Ved å neglisjere kohorteffekten ser vi (med små avvik) at premiene blir underestimert. Over lengre tid kan dette ha store økonomiske konsekvenser med tanke på utgifter for forsikringsselskaper. Dette stemmer også med tidligere resultater der kohorteffekten fører til litt lavere dødeligheter for både menn og kvinner. Det skal riktignok sies at forskjellen kun blir betydelig for yngre kohorter. Tilsvarende resultater får vi også for svensk- og danskbefolkning. Analysen for disse ligger i Appendix A, og er kun gjennomført for kvinner. Disse bygger opp under påstander og resultater vi har kommet frem til i oppgaven, og viser at modellen også fungerer for andre land i Skandinavia.

Oppsummert, gjennom resultater og analyser i oppgaven, er Gompertz-Makeham en fin modell for estimering av kohortdata. Modellen er over 150 år gammel og brukes fortsatt mest flittig for dødeligheter. Små parametriske formuleringer måtte til for at modellen tilpasset seg kohortdødeligheten godt. Det var et mål å komme frem til enkle dynamiske formuleringer og opplegg for fremskrivninger under slike kohort-baserte modeller, noe som ble oppnådd i denne oppgaven. Modellering gjennom algoritmer har påvist endringer i resultater mellom kohort- og periodedata, og de økonomiske forskjeller som oppstod når kohortinformasjonen i dataene ikke ble oversett. Dynamisk modellering av kohorter har både sine styrker og svakheter, men er essensielt viktig når det kommer til økonomistyring.

6.1 Svakheter med modellen

Det ble underveis i oppgaven redegjort for at et vanlig problem med Gompertz-Makeham modellen er at den forlanger økende dødelighetssannsynligheter. Dette var et stort problem for oss da mange av kohortene ikke var fullstendige. Dette førte til noen restriksjoner for å omgå problemet. En direkte følge av dette var at vi for noen aldre underestimerte premiene. Det ble heller ikke inkludert inflasjon i diskonteringsfaktoren, noe som også kan endre premiene.

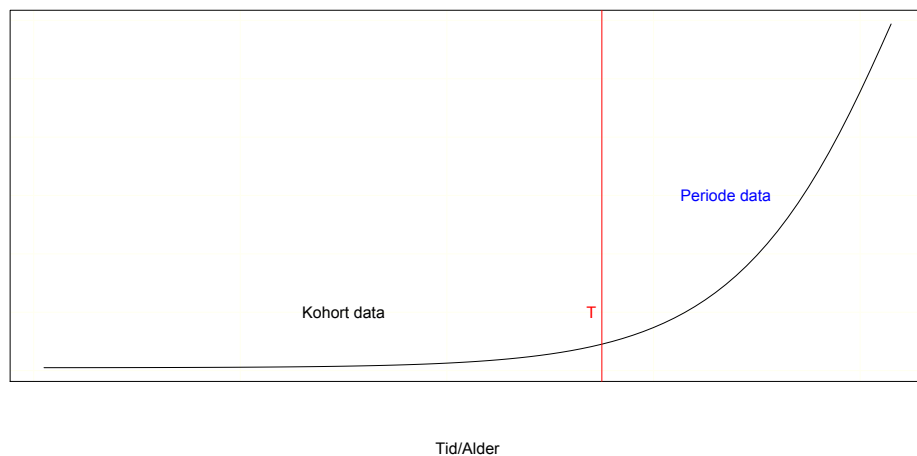
Det største problemet var mangelen på fullstendig datasett. Kohortdødelighet bygger på at individer i de ulike kohortene har dødd ut, slik at gjennomsnittet av dødeligheten over alle individene betegner dødeligheten for den gitte kohorten. Naturlig nok vil ikke de yngre kohortene ha kohortdødelighet for eldste aldre, og simuleringer måtte til som erstattet disse. Dette er et problem som gikk gjennom hele oppgaven, og som også ledet til restriksjoner forklart ovenfor. Kohorteffekten ble kun betydelig for de yngre kohortene, og dette kan rett og slett ha med datamangler å gjøre. I det økonomiske eksemplet var differansen i premieverdien størst for de eldste aldre, altså for de yngre kohortene.

6.2 Fremtidig arbeid

I avsnitt 2.4 ble det kort forklart om tidligere arbeid gjort på kohortestimeringer. Mange tilnærminger er gjort for effekter på tvers av alder. I tillegg til Gompertz-Makeham er Lee-Carter også en velkjent og velbrukt modell for kohortdødeligheter. Forskere og aktuarer den dag i dag jobber med å finne en ultimat og optimal modell som skal funke over alle type kohortdata. Til nå er det mange som utvider eksisterende modeller til å passe dataene, og det ser ut til at forskjellige type datasett krever forskjellige type utvidelser.

Forslag til utbedring av modellen testet i denne oppgaven vil være å utvide modellen til å ta hensyn til minkende dødelighetssannsynligheter for de eldre aldre. Det mest optimale vil være å modifisere modellen til å kunne behandle ikke-eksisterende data for visse kohorter og aldre. Ved å ikke sette restriksjoner, men heller la modellen naturlig behandle nulldata vil trolig føre til mindre underestimering. For støyleddene i Gompertz-Makehams modell ble en normalantagelse brukt, noe som kan være en svakhet ved modellen da det er mulig det finnes bedre fordelinger som beskriver støyen bedre.

Det største problemet med kohortmodellering virker som sagt til å være mangel på data. En mulig ide for å fange opp kohorteffekt i data, og likevel ha dødelighetssannsynligheter for de eldste aldre kan være som følger:

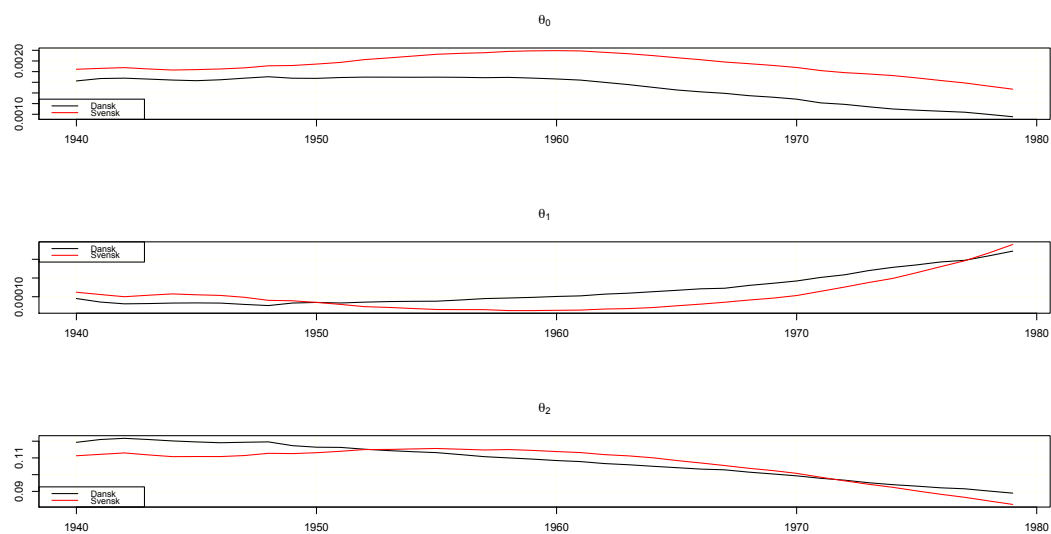


Frem til et visst tidspunkt T , der det for eksempel begynner å bli mangel på kohortdata, kan det være en mulighet å bytte over til periode-dødelighetsdata. Dette betyr at vi vil ha et datasett som både består av kohortdødeligheter opptil et tidspunkt T , og periodedødelighet videre. Da kan det være mulig å simulere effektene av individene i kohorter inn i periodedødelighetene etter T , slik at vi bevarer kohortinformasjonen. Dette må bli gjort på en slik måte at periodedødeligheten ikke er et estimat for den fremtidige kohortdødeligheten (avsnitt 2.3), men heller bygger opp under den. Dette er kun en mulig fremgangsmåte.

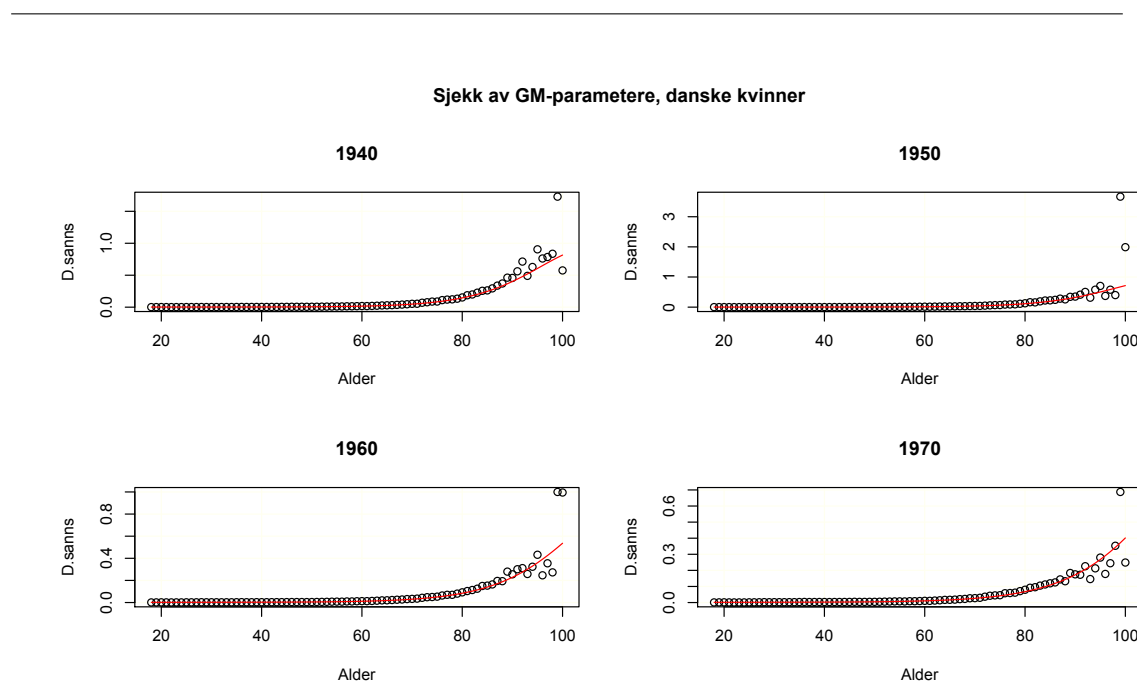
7 Appendix

A: Resultater for svensk og dansk befolkning

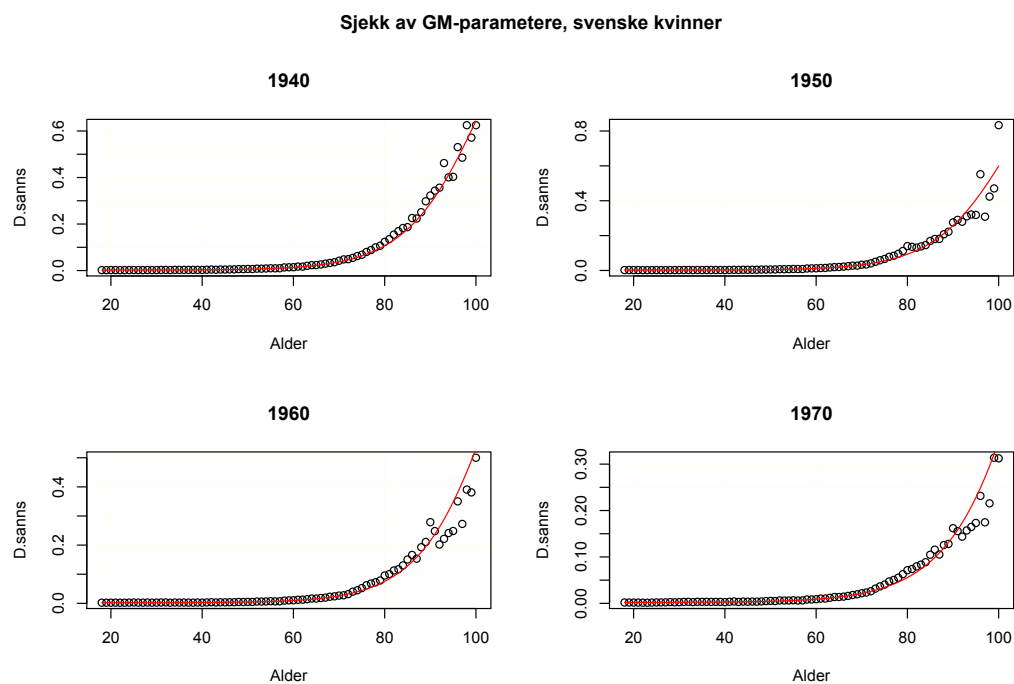
Påfølgende resultater er kun med hensyn på danske og svenske kvinner. Det er plukket ut utvalgte kohorter for visse aldre og år, men mye av resultatene svarer til samme dimensjon og lengder som ved den norske befolkningen.



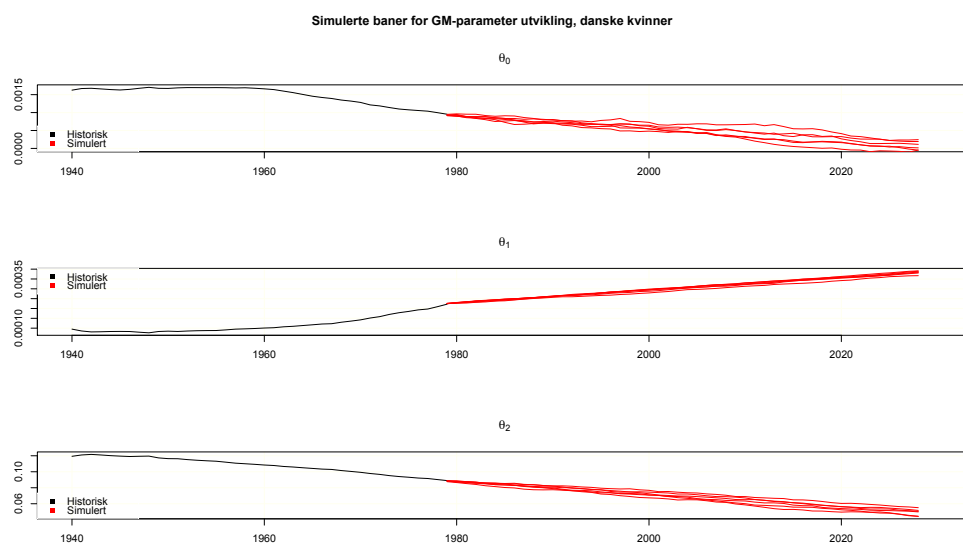
Figur A1: *Parametere Gompertz-Makeham modell*



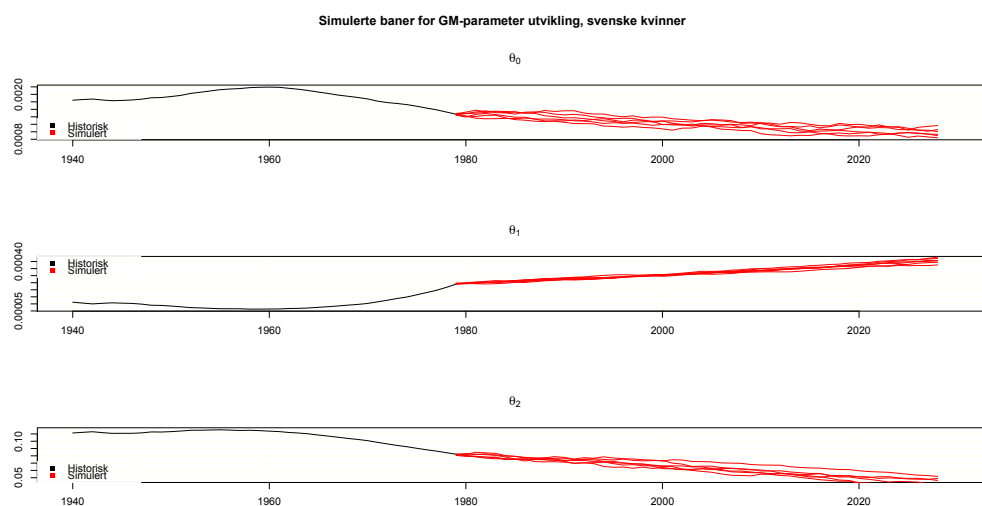
Figur A2: *Rådata mot Gompertz-Makeham modell*



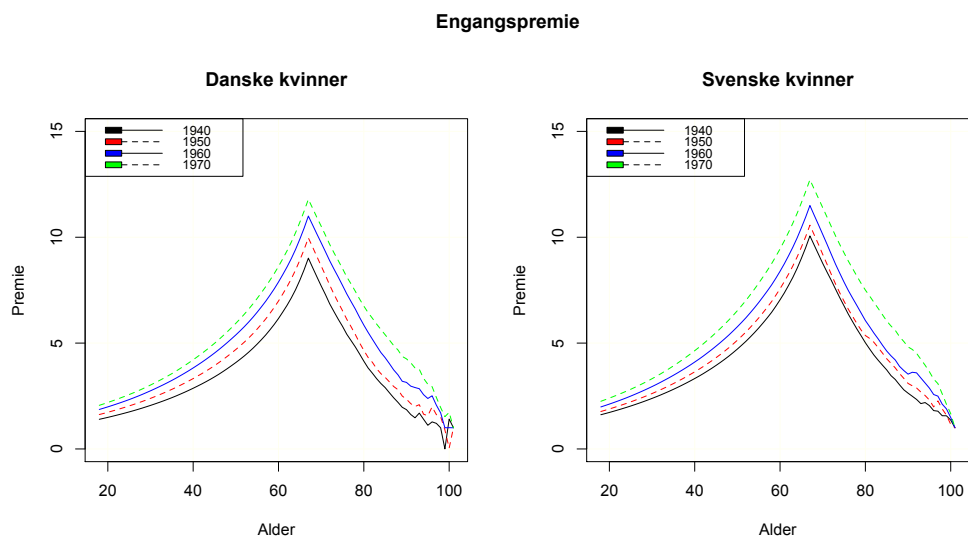
Figur A3: *Rådata mot Gompertz-Makeham modell*



Figur A4: *Simulerte baner for utviklingen av GM-parametere*



Figur A5: *Simulerte baner for utviklingen av GM-parametere*



Figur A5: *Utvalg av premie med pensjonsalder 67*

Premie						
	1940			1950		
	30	60	90	30	60	90
Danske kvinner	2/ 2.2	6.1/ 6.5	1.9/ 2	2.4/ 2.6	7/ 7.6	2.3/ 2
Svenske kvinner	2.3/ 2.5	7/ 7.3	2.6/ 2.5	2.6/ 2.8	7.6/ 8.1	3.1/ 3
Norske kvinner	2.5/ 2.7	8/ 8.4	4.2/ 4.1	2.9/ 3	8.7/ 9.2	5/ 4.5
	1960			1970		
	30	60	90	30	60	90
Danske kvinner	2.7/ 3.6	7.9/ 8.9	3.1/ 3	3/ 4.1	8.6/ 9.7	4.2/ 4.6
Svenske kvinner	2.9/ 3.5	8.4/ 9.5	3.5/ 3.4	3.3/ 4.4	9.4/ 10.7	4.8/ 5.1
Norske kvinner	3.2/ 3.8	9.6/ 10.6	5.6/ 5.4	3.6/ 4.6	10.7/ 12	6.3/ 6.6

Tabell A1: *Premie-beregninger for danske, svenske og norske kvinner, med og uten kohorteffekt. Som før, uten kohorteffekt i **rødt***

B: Tabeller

Tabell B1: Minus log-likelihood estimator

År	Menn	Kvinner
1940	154782.3498	124051.088
1941	120944.5496	94003.760
1942	130517.2307	101209.941
1943	121391.9839	91533.797
1944	122697.8305	82845.867
1945	101075.8631	71032.978
1946	93741.2603	58722.057
1947	63751.1201	32020.697
1948	49532.8561	18339.170
1949	31160.4399	-4592.361
1950	13230.1220	-18111.886
1951	9937.3110	-22195.578
1952	105.7807	-32056.420
1953	-11627.2575	-40896.939
1954	-18196.9899	-51130.432
1955	-24371.4713	-57407.843
1956	-33336.2186	-61620.009
1957	-38577.2956	-70659.351
1958	-45012.8662	-76900.217
1959	-50190.5286	-81376.568
1960	-56237.4357	-86865.351
1961	-61463.9765	-91567.230
1962	-65159.3057	-97736.795
1963	-72585.7034	-110618.906
1964	-68734.3246	-100023.945
1965	-70970.2278	-100654.031
1966	-77871.6228	-109513.770
1967	-77961.8004	-108768.688
1968	-81205.5398	-115621.643
1969	-86418.0838	-117827.453
1970	-89861.6984	-123823.665
1971	-90022.2214	-131656.217
1972	-93863.5575	-133437.677
1973	-96168.5845	-139932.388
1974	-96798.9094	-138974.317
1975	-99921.5263	-143926.288
1976	-104220.5151	-148369.287
1977	-103593.3031	-147575.878
1978	-106188.2946	-155142.321
1979	-111014.9494	-165795.310

Tabell B2: Gompertz Makeham parameter-estimer fra kohortdata, menn

År	θ_0	θ_1	θ_2
1940	0.004918834	1.103365e-04	0.1097808
1941	0.004862515	1.067303e-04	0.1100459
1942	0.004813177	9.999105e-05	0.1109830
1943	0.004783120	9.796634e-05	0.1109057
1944	0.004769249	9.537974e-05	0.1109398
1945	0.004743884	9.475044e-05	0.1105466
1946	0.004748517	9.102819e-05	0.1107483
1947	0.004777817	8.510879e-05	0.1113337
1948	0.004798865	7.839997e-05	0.1123218
1949	0.004825321	7.365164e-05	0.1128934
1950	0.004841487	7.008969e-05	0.1132897
1951	0.004878464	6.360350e-05	0.1145893
1952	0.004898849	6.148697e-05	0.1145420
1953	0.004931052	5.887671e-05	0.1147123
1954	0.004972580	5.379682e-05	0.1157498
1955	0.005002401	5.147273e-05	0.1159293
1956	0.005025396	4.743286e-05	0.1168829
1957	0.005089426	4.314721e-05	0.1178875
1958	0.005118719	4.044849e-05	0.1184516
1959	0.005141860	3.834074e-05	0.1188664
1960	0.005181087	3.544324e-05	0.1196933
1961	0.005205962	3.362342e-05	0.1201220
1962	0.005195615	3.193090e-05	0.1206003
1963	0.005158572	3.084705e-05	0.1208509
1964	0.005119077	2.966085e-05	0.1212503
1965	0.005068907	2.934613e-05	0.1210593
1966	0.005017441	2.968996e-05	0.1204814
1967	0.004981174	2.920430e-05	0.1204909
1968	0.004944383	2.908396e-05	0.1201823
1969	0.004907082	2.946782e-05	0.1195954
1970	0.004874191	2.968892e-05	0.1191615
1971	0.004857211	2.936944e-05	0.1190412
1972	0.004812792	2.987773e-05	0.1184628
1973	0.004779077	3.019575e-05	0.1179564
1974	0.004735522	3.062392e-05	0.1174406
1975	0.004693962	3.170934e-05	0.1165245
1976	0.004645442	3.313387e-05	0.1155007
1977	0.004610460	3.366184e-05	0.1149669
1978	0.004569968	3.519384e-05	0.1138865
1979	0.004519458	3.769288e-05	0.1124602

Tabell B3: Gompertz Makeham parameter-estimer fra kohortdata, kvinner

År	θ_0	θ_1	θ_2
1940	0.004721206	1.524423e-04	0.09878049
1941	0.004673923	1.466602e-04	0.09921272
1942	0.004655796	1.383279e-04	0.10000956
1943	0.004620417	1.386214e-04	0.09954550
1944	0.004597045	1.380262e-04	0.09912597
1945	0.004599188	1.327153e-04	0.09937098
1946	0.004595865	1.305376e-04	0.09914481
1947	0.004657980	1.253039e-04	0.09944150
1948	0.004692720	1.162904e-04	0.10032378
1949	0.004713526	1.115479e-04	0.10058538
1950	0.004753367	1.039970e-04	0.10140123
1951	0.004789460	9.612456e-05	0.10239387
1952	0.004842774	9.130425e-05	0.10271713
1953	0.004899954	8.517033e-05	0.10337970
1954	0.004956039	7.886239e-05	0.10415714
1955	0.005012969	7.264158e-05	0.10505511
1956	0.005057150	6.670986e-05	0.10600413
1957	0.005114451	6.226672e-05	0.10664056
1958	0.005149950	5.835035e-05	0.10725535
1959	0.005185544	5.450677e-05	0.10792667
1960	0.005237447	5.056321e-05	0.10867823
1961	0.005257247	4.794424e-05	0.10908104
1962	0.005262969	4.545183e-05	0.10950872
1963	0.005245993	4.313783e-05	0.10995676
1964	0.005223188	4.049371e-05	0.11072616
1965	0.005171375	4.111050e-05	0.11003050
1966	0.005132992	4.163226e-05	0.10929640
1967	0.005115186	4.028516e-05	0.10925816
1968	0.005095413	4.061528e-05	0.10856642
1969	0.005059925	4.164802e-05	0.10766702
1970	0.005030275	4.362298e-05	0.10637821
1971	0.005010437	4.506092e-05	0.10533305
1972	0.004980133	4.571808e-05	0.10465059
1973	0.004945485	4.743577e-05	0.10354549
1974	0.004907628	4.919313e-05	0.10252859
1975	0.004856931	5.312692e-05	0.10078797
1976	0.004807958	5.718496e-05	0.09912727
1977	0.004759281	6.151358e-05	0.09745483
1978	0.004689847	6.967883e-05	0.09493138
1979	0.004627605	7.835645e-05	0.09254881

Tabell B4: Gompertz Makeham parameter-estimer fra periodedata, menn

År	θ_0	θ_1	θ_2
1940	0.004882290	1.096607e-04	0.10972010
1941	0.004832908	1.043043e-04	0.11033139
1942	0.004794570	1.001686e-04	0.11072538
1943	0.004769893	9.781351e-05	0.11070056
1944	0.004755053	9.584246e-05	0.11058477
1945	0.004740840	9.358879e-05	0.11049713
1946	0.004756718	8.872959e-05	0.11088468
1947	0.004783463	8.266923e-05	0.11159927
1948	0.004806155	7.665475e-05	0.11244955
1949	0.004828757	7.258871e-05	0.11291052
1950	0.004853532	6.759213e-05	0.11371081
1951	0.004885847	6.335106e-05	0.11433396
1952	0.004911214	6.064486e-05	0.11447774
1953	0.004947857	5.681419e-05	0.11505268
1954	0.004979601	5.337109e-05	0.11558196
1955	0.005006229	5.016117e-05	0.11614393
1956	0.005052524	4.571360e-05	0.11719296
1957	0.005098976	4.215763e-05	0.11799308
1958	0.005126364	3.975864e-05	0.11849349
1959	0.005158862	3.721988e-05	0.11910994
1960	0.005194379	3.474357e-05	0.11977133
1961	0.005194194	3.312103e-05	0.12017273
1962	0.005172602	3.170888e-05	0.12053680
1963	0.005132569	3.068429e-05	0.12078947
1964	0.005084219	2.992783e-05	0.12090241
1965	0.005038050	2.990029e-05	0.12054076
1966	0.004995107	2.988164e-05	0.12022205
1967	0.004957557	2.953086e-05	0.12010074
1968	0.004922017	2.963696e-05	0.11965996
1969	0.004889825	2.990607e-05	0.11916713
1970	0.004864952	2.982419e-05	0.11891571
1971	0.004841361	2.976079e-05	0.11863596
1972	0.004790280	3.032957e-05	0.11802195
1973	0.004750137	3.099968e-05	0.11736787
1974	0.004709800	3.164322e-05	0.11671515
1975	0.004663950	3.293915e-05	0.11573547
1976	0.004627754	3.381225e-05	0.11499818
1977	0.004586012	3.481982e-05	0.11421724
1978	0.004540247	3.687165e-05	0.11295065
1979	0.004496110	3.896801e-05	0.11175375

1980	0.004446769	4.196577e-05	0.11023672
1981	0.004392145	4.503105e-05	0.10882245
1982	0.004333842	4.814413e-05	0.10748539
1983	0.004267910	5.169359e-05	0.10606982
1984	0.004202643	5.533619e-05	0.10473685
1985	0.004147172	5.813454e-05	0.10375605
1986	0.004065944	6.354199e-05	0.10210077
1987	0.003969946	7.121038e-05	0.10001069
1988	0.003860613	8.080623e-05	0.09773550
1989	0.003786946	8.976248e-05	0.09575963
1990	0.003733630	9.696202e-05	0.09427712
1991	0.003663222	1.074174e-04	0.09234714
1992	0.003603466	1.148229e-04	0.09102844
1993	0.003568916	1.166743e-04	0.09064448
1994	0.003557122	1.159879e-04	0.09058395
1995	0.003550859	1.161976e-04	0.09037063
1996	0.003544829	1.166542e-04	0.09010630
1997	0.003542051	1.157707e-04	0.09002436
1998	0.003526549	1.165278e-04	0.08969728
1999	0.003513279	1.153463e-04	0.08966611
2000	0.003500451	1.141478e-04	0.08964390
2001	0.003493471	1.134829e-04	0.08948164
2002	0.003502114	1.100798e-04	0.08973710
2003	0.003502534	1.075485e-04	0.08985550
2004	0.003497660	1.067475e-04	0.08963358
2005	0.003494496	1.053300e-04	0.08948649
2006	0.003464645	1.046556e-04	0.08923696
2007	0.003421587	1.023516e-04	0.08926955
2008	0.003356166	1.006849e-04	0.08925184
2009	0.003306303	9.817431e-05	0.08933412

Tabell B5: Gompertz Makeham parameter-estimer fra periodedata, kvinner

År	θ_0	θ_1	θ_2
1940	0.004696163	1.502755e-04	0.09889234
1941	0.004657457	1.440568e-04	0.09941378
1942	0.004634790	1.398345e-04	0.09958945
1943	0.004608249	1.390397e-04	0.09923257
1944	0.004593558	1.362107e-04	0.09911263
1945	0.004596018	1.323774e-04	0.09914501
1946	0.004617831	1.288626e-04	0.09915791
1947	0.004667819	1.219450e-04	0.09969908
1948	0.004707860	1.142688e-04	0.10037289
1949	0.004728553	1.086534e-04	0.10083366
1950	0.004771222	1.007768e-04	0.10174965
1951	0.004817520	9.425249e-05	0.10242495
1952	0.004874668	8.802125e-05	0.10309009
1953	0.004921280	8.281046e-05	0.10358953
1954	0.004981033	7.640171e-05	0.10443071
1955	0.005024510	7.045739e-05	0.10533865
1956	0.005082869	6.493155e-05	0.10618384
1957	0.005131205	6.101324e-05	0.10673539
1958	0.005163664	5.707303e-05	0.10737459
1959	0.005210198	5.284108e-05	0.10817206
1960	0.005243261	4.983419e-05	0.10866758
1961	0.005225161	4.898763e-05	0.10852983
1962	0.005250388	4.466189e-05	0.10957541
1963	0.005232514	4.212682e-05	0.11018751
1964	0.005193721	4.105268e-05	0.11026106
1965	0.005145750	4.190319e-05	0.10943522
1966	0.005125809	4.136433e-05	0.10909057
1967	0.005102571	4.096162e-05	0.10868496
1968	0.005073264	4.176899e-05	0.10784826
1969	0.005039581	4.311988e-05	0.10681675
1970	0.005019811	4.477932e-05	0.10567307
1971	0.004989290	4.600362e-05	0.10475928
1972	0.004957087	4.735896e-05	0.10382695
1973	0.004919905	4.910735e-05	0.10277285
1974	0.004875679	5.154691e-05	0.10150332
1975	0.004828023	5.594940e-05	0.09969866
1976	0.004778004	6.019025e-05	0.09802085
1977	0.004725979	6.612801e-05	0.09601401
1978	0.004655460	7.480653e-05	0.09351980
1979	0.004595652	8.318875e-05	0.09134134

1980	0.004528309	9.348121e-05	0.08902581
1981	0.004453531	1.058320e-04	0.08657079
1982	0.004357726	1.216390e-04	0.08390232
1983	0.004253512	1.429891e-04	0.08086060
1984	0.004149655	1.643092e-04	0.07825310
1985	0.004052474	1.866785e-04	0.07586316
1986	0.003930286	2.151802e-04	0.07327184
1987	0.003773754	2.584031e-04	0.06998196
1988	0.003632180	3.001324e-04	0.06731589
1989	0.003526671	3.361686e-04	0.06523251
1990	0.003420278	3.761671e-04	0.06317582
1991	0.003303098	4.199357e-04	0.06119915
1992	0.003211053	4.537129e-04	0.05973525
1993	0.003148665	4.758867e-04	0.05882512
1994	0.003106825	4.873108e-04	0.05828556
1995	0.003104601	4.917160e-04	0.05790802
1996	0.003076158	5.022692e-04	0.05740003
1997	0.003058171	5.091612e-04	0.05698762
1998	0.003054066	5.107123e-04	0.05674285
1999	0.003037440	5.120589e-04	0.05655050
2000	0.003034444	5.053369e-04	0.05661686
2001	0.003034888	5.055553e-04	0.05643121
2002	0.003050018	4.920207e-04	0.05672498
2003	0.003045204	4.869815e-04	0.05672836
2004	0.003034808	4.862634e-04	0.05655888
2005	0.003032074	4.823932e-04	0.05645728
2006	0.003017016	4.776599e-04	0.05641320
2007	0.002989612	4.706795e-04	0.05650135
2008	0.002939575	4.678050e-04	0.05646729
2009	0.002888741	4.654748e-04	0.05637379

Tabell B6: Gompertz Makeham parameter-estimates for danske kvinder

År	θ_0	θ_1	θ_2
1940	0.00162536688	0.00009515242	0.119318467
1941	0.00167041222	0.00008574241	0.121007528
1942	0.00167749776	0.00008097493	0.121705676
1943	0.00166004365	0.00008173070	0.120999825
1944	0.00164185497	0.00008305982	0.120155595
1945	0.00163014243	0.00008346121	0.119558632
1946	0.00164730733	0.00008300548	0.119068584
1947	0.00167739010	0.00007942680	0.119360318
1948	0.00170410169	0.00007638128	0.119581641
1949	0.00167641533	0.00008308497	0.117330089
1950	0.00167362921	0.00008477951	0.116409869
1951	0.00168935989	0.00008323978	0.116291688
1952	0.00169689291	0.00008549788	0.115206418
1953	0.00169619150	0.00008699995	0.114384058
1954	0.00169443465	0.00008773830	0.113724112
1955	0.00169616459	0.00008805152	0.113165636
1956	0.00169344227	0.00009128306	0.111983252
1957	0.00168635436	0.00009504385	0.110746835
1958	0.00169202312	0.00009655575	0.110012535
1959	0.00167856702	0.00009828640	0.109290399
1960	0.00166190231	0.00010069521	0.108431601
1961	0.00164098824	0.00010230593	0.107810271
1962	0.00159860573	0.00010689500	0.106628511
1963	0.00155633953	0.00010951175	0.105925244
1964	0.00150597567	0.00011335798	0.105015593
1965	0.00145618173	0.00011711447	0.104180079
1966	0.00142096867	0.00012114384	0.103327414
1967	0.00139285318	0.00012266160	0.102893391
1968	0.00134930625	0.00013013808	0.101516057
1969	0.00132062545	0.00013585360	0.100423679
1970	0.00128338843	0.00014211199	0.099298994
1971	0.00121407347	0.00015155328	0.097823277
1972	0.00118601528	0.00015859032	0.096699352
1973	0.00114070353	0.00016986746	0.095178624
1974	0.00109921937	0.00017851530	0.094020488
1975	0.00107660867	0.00018506082	0.093151828
1976	0.00105748675	0.00019292655	0.092142289
1977	0.00103881081	0.00019734783	0.091558099
1978	0.00099764656	0.00020915635	0.090232521
1979	0.00095389803	0.00022186758	0.088960347

Tabell B7: Gompertz Makeham parameter-estimator for svenske kvinner

År	θ_0	θ_1	θ_2
1940	0.00184416941	0.00011200620	0.111285032
1941	0.00186087599	0.00010584928	0.112146999
1942	0.00187478977	0.00010000880	0.112978060
1943	0.00185006738	0.00010399361	0.111800872
1944	0.00182994620	0.00010750691	0.110765468
1945	0.00183800518	0.00010518715	0.110842975
1946	0.00184947736	0.00010347191	0.110812273
1947	0.00187067625	0.00009846987	0.111419650
1948	0.00190907312	0.00009030962	0.112753462
1949	0.00191489188	0.00008913289	0.112589375
1950	0.00194160237	0.00008482590	0.113123578
1951	0.00197275701	0.00007947348	0.113963260
1952	0.00202565428	0.00007348367	0.114998820
1953	0.00205742335	0.00007140227	0.115067755
1954	0.00209105054	0.00006852919	0.115367719
1955	0.00212532535	0.00006584847	0.115641844
1956	0.00214274529	0.00006554881	0.115206958
1957	0.00215482168	0.00006547088	0.114746737
1958	0.00217927332	0.00006296599	0.114997020
1959	0.00218947766	0.00006298209	0.114500734
1960	0.00219360113	0.00006367289	0.113796632
1961	0.00218826946	0.00006427647	0.113192546
1962	0.00216163198	0.00006708272	0.111964559
1963	0.00213556619	0.00006830598	0.111225880
1964	0.00210120998	0.00007100046	0.110094259
1965	0.00206116027	0.00007575986	0.108455245
1966	0.00202375242	0.00008018598	0.106995685
1967	0.00198038914	0.00008523959	0.105453758
1968	0.00194931806	0.00009108514	0.103814429
1969	0.00191673754	0.00009615914	0.102448402
1970	0.00187751164	0.00010315309	0.100786823
1971	0.00181995665	0.00011444824	0.098469204
1972	0.00178082567	0.00012580873	0.096312918
1973	0.00175610461	0.00013796335	0.094244171
1974	0.00172633281	0.00014885240	0.092454244
1975	0.00168211297	0.00016427091	0.090298811
1976	0.00163373798	0.00018035908	0.088275774
1977	0.00158770180	0.00019602944	0.086477294
1978	0.00152842479	0.00021684287	0.084321466
1979	0.00147110462	0.00023961286	0.082217223

C: Programkode i R

Utelater programkode for grafisk fremstilling/plott av resultater

Programkode for innlesning og sortering av data

```
#kohort dødelighetssannsynligheter menn
df <-read.table("/Users/Jaz/Masteroppgave/R-koder/NOR/STATS/
  mltcoh_1x1.txt",header=TRUE,skip = 2, na.strings=c(".", "0"), sep="")

#trekker ut alle menn i aldre 18-100
coh_MA= df[df$Age %in% 18:100,]

#kohort dødelighetssannsynligheter kvinner
df1 <-read.table("/Users/Jaz/Masteroppgave/R-koder/NOR/STATS/
  fltcoh_1x1.txt",header=TRUE,skip = 2, na.strings=c(".", "0"), sep="")

#trekker ut alle kvinner i aldre 18-100
coh_FE = df1[df1$Age %in% 18:100,]

# utsatt for risiko for å dø, kohort
exp <-read.table("/Users/Jaz/Masteroppgave/R-koder/NOR/STATS/
  cExposures_1x1.txt",header=TRUE,skip = 2,
  na.strings=c(".", "0"), sep="")
#antall individer utsatt for risiko for å dø
exposure = exp[exp$Age %in% 18:100,] #henter ut for aldre 18-100
exp_RISK = exposure[exposure$Year %in% 1940:1979,] #for år 1940++
#exp_RISK[is.na(exp_RISK)] = 0

#antall døde individer for kohorter
d <-read.table("/Users/Jaz/Masteroppgave/R-koder/NOR/STATS/
  cDeaths_1x1.txt",header=TRUE,skip = 2,
  na.strings=c(".", "0"), sep="")
#alle dødsfall
dead = d[d$Age %in% 18:100,] #alle dødsfall for år 18-100
deaths = dead[dead$Year %in% 1940:1979,] #for år 1940++

#setter alle relevante data for funksjon i et datasett
set.seed(1)
risk = expand.grid(Age=18:100, Year = 1940:1979)
risk$Female = exp_RISK$Female
risk$Male = exp_RISK$Male
risk$Total = exp_RISK$Total
```

```

dead = expand.grid(Age=18:100,Year=1940:1979)
dead$Female = deaths$Female
dead$Male = deaths$Male
dead$Total = deaths$Total

# sammenslår til et datasett
f_coh = merge(risk,dead,by=c("Year","Age"),
              suffixes=c(".risk",".death"))

#antall individer utsatt for risiko for å dø periodedata
p <-read.table("/Users/Jaz/Masteroppgave/R-koder/NOR/STATS/
              Exposures_1x1.txt",header=TRUE,skip = 2,
              na.strings=c(".", "0"), sep="")

E = p[p$Age %in% 18:100,] #henter ut for aldre 18-100
periode = E[E$Year %in% 1940:2009,] #for år 1940++

#antall dødsfall for periode data
d <-read.table("/Users/Jaz/Masteroppgave/R-koder/NOR/STATS/
              Deaths_1x1.txt",header=TRUE,skip = 2,
              na.strings=c(".", "0"), sep="")

dead = d[d$Age %in% 18:100,] #alle dødsfall for år 18-100
deaths = dead[dead$Year %in% 1940:2009,] #for år 1940++

#sammenslår til et datasett for periode data
periode = data.frame(periode$Year,periode$Age,periode$Male,
                    periode$Female, deaths$Male,deaths$Female)
colnames(periode) = c("Year","Age","Male.risk","Female.risk",
                    "Male.death","Female.death")

```

Programkode for funksjoner brukt gjennom oppgaven

```

#funksjon for å legge bakgrunnsfarge på figurer

ggbg= function() {
  points(0,0,pch=16,cex=1e6,col="ivory")
  grid(col="ivory",lty=1)
}

```

```

}

#estimerer log-likelihood funksjon under GM-modell

minusgompakll1=function(s,ll,n,z){
  t=exp(s)
  p=exp(-t[1]-t[2]*exp(t[3]*ll))
  -sum((n-z)*log(p)+z*log(1-p))
}

#likelihood estimerer av GM-parametere blir estimert

gompakfit=function(ll,n,z){
  s=-c(8,9,2)
  o=optim(s,minusgompakll1,"BFGS",ll=ll,n=n,z=z)
  theta=exp(o$par)
  list(theta=theta)
}

#returnerer forventet gjenværende leveår i alder "l_0" under GM-modell
#med parameter "theta"

gompakmean=function(theta,y_0,h){
  l_0=y_0/h
  l_e=110/h
  ll=(l_0+1):l_e
  hh=(theta[2]/theta[3])*(exp(theta[3]*h)-1)
  p=exp(-theta[1]*h-hh*exp(theta[3]*ll*h))
  P=cumprod(p)
  (0.5+sum(P))*h
}

#livstabell blir konstruert "K" perioder fremover for hver startalder.
#gir ut livstabell "kp" med en kolonne per startalder

lifetab=function(q,K){
  l_e=length(q)
  p=c(1-q,rep(0,l_e+1))
  kp=matrix(1,K,l_e+1)
  for(l in 0:l_e+1) kp[,l] = cumprod(p[1:K+(l-1)])
  kp=rbind(rep(1,l_e+1),kp)
  list(kp=kp)
}

```

```

#estimerer engangspremie for hver alder i dødelighetsvektor "q".
#"s" er pensjonen fra alder "l_r", og "d" er diskonteringsfaktor.
#gir ut engangspremie "pi" (vektor)

onetimeprem=function(q,s,l_r,d){
  #først livstabell
  l_e=length(q)
  kp=lifetab(q,l_e)$kp
  # så indikator matrise
  I=matrix(0,l_e+1,l_e+1)
  I[row(I)+col(I)>l_r+1]=1
  #beregner engangspremie
  ll=0:l_e
  M=s*(d**ll)*kp*I
  pi=apply(M,2,sum)
  list(pi=pi)
}

```

Programkode for innledende datastudier

```

library("demography")

#prosjeksjoner av dødelighetsrater

nor_demo = read.demogdata("/Users/Jaz/Masteroppgave/R-koder/
                          NOR/STATS/cMx_1x1.txt", type="mortality",
                          skip=2,label="Norway")

par(mfrow=c(1,2))
plot(nordemo, series="Male", datatype="rate", xlab="Alder",
      ylab="", main = "Rater menn")
plot(nordemo,series="Female",datatype="rate", xlab="Alder",
      ylab="", main="Rater kvinner")

par(mfrow=c(1,2))
plot(nordemo,series="Male",datatype="rate", plot.type="time",
      main="Menn",xlab="År",ylab="")
plot(nordemo,series="Female",datatype="rate",plot.type="time",
      main="Kvinner",xlab="År",ylab="")

```

```

#yngre kohorter

i=c(1940:1979)
ii = c(18:100)
#Som ovenfor med "i" og "ii" inkludert

#kjører funksjon over alle år i datasett.
#får ut minus log-likelihood verdier

s=-c(8,9,2)
library(dplyr)
ming_menn = f_coh%>%
  group_by(Year) %>%
  summarise(likelihood=minusgompakll1(s, Age,
    Male.risk, Male.death))
  #for menn
ming_kvinner = f_coh%>%
  group_by(Year) %>%
  summarise(likelihood=minusgompakll1(s, Age,
    Female.risk, Female.death))
  #for kvinner

```

Gompertz-Makeham parametere og analyse for kohortdata

```

#konstruerer matriser og estimerer parametere

aM = matrix(NA, nrow=(100-18+1), ncol=40)
aK = matrix(NA, nrow=(100-18+1), ncol=40)
dM = matrix(NA, nrow=(100-18+1), ncol=40)
dK = matrix(NA, nrow=(100-18+1), ncol=40)
qM =matrix(NA, nrow=(100-18+1), ncol=40)
qK =matrix(NA, nrow=(100-18+1), ncol=40)

t = 1
for(i in 1:(100-18+1)){
  for(j in 1:40){
    aM[i,j]=(f_coh[f_coh$Year %in% (j+1939),])$Male[t]
    aK[i,j]=(f_coh[f_coh$Year %in% (j+1939),])$Female[t]
    dM[i,j]=f_coh$Male.death[t]

```

```

        dK[i,j]=f_coh$Male.death[t]
    }
    t=t+1
}
#får ut dødelighetssannsynligheter
qM = dM/aM
qK = dK/aK

#matriser for menn og kvinner GM-parametere
theta_menn = matrix(NA,nrow=40, ncol=3)
theta_kvinner = matrix(NA,nrow=40, ncol=3)
s=-c(8,9,2) #startverdier

#kjører over dataene og estimerer GM-parametere
for(j in 1:40){
    thm = optim(-c(8,9,2),minusgompakll1,"BFGS",ll=(1:83),
               n=aM[,j],z=dM[,j])
    thk = optim(-c(8,9,2),minusgompakll1,"BFGS",ll=(1:83),
               n=aK[,j],z=dM[,j])

    theta_menn[j,] = exp(thm$par)
    theta_kvinner[j,] = exp(thk$par)
}

#teste om modell og parametere passer rådata

q_x_m = matrix(NA, nrow=(100-18+1), ncol=40)
q_x_k = matrix(NA, nrow=(100-18+1), ncol=40)
x = (0:82)

for(i in 1:(100-18+1)){
    for(j in 1:40){
        q_x_m[i,j] = 1-exp(-theta_menn[j,1]-theta_menn[j,2]*
                           exp(theta_menn[j,3]*x[i]))
        q_x_k[i,j] = 1-exp(-theta_kvinner[j,1]-theta_kvinner
                           [j,2]*exp(theta_kvinner[j,3]*x[i]))
    }
}

#plott forventede leveår med estimerte theta-parametere for
#utvalgte år
a = 18
b = 100

```

```

E_menn=rep(0,b+1)
E_kvinner=rep(0,b+1)
for(i in a:b){
  E_menn[i+1]=gompakmean(theta_menn["år",],i,h=1)
  E_kvinner[i+1]=gompakmean(theta_kvinner["år",],i,h=1)
}

```

Gompertz-Makeham parametere og analyse for periodedata

```

#konstruerer matriser for å beregne dødelighet

pM = matrix(NA, nrow=(100-18+1), ncol=70)
pK = matrix(NA, nrow=(100-18+1), ncol=70)
dM = matrix(NA, nrow=(100-18+1), ncol=70)
dK = matrix(NA, nrow=(100-18+1), ncol=70)
qM_p =matrix(NA, nrow=(100-18+1), ncol=70)
qK_p =matrix(NA, nrow=(100-18+1), ncol=70)

t = 1
for(i in 1:(100-18+1)){
  for(j in 1:70){
    pM[i,j]=(periode[periode$Year %in% (j+1939),])$Male[t]
    pK[i,j]=(periode[periode$Year %in% (j+1939),])$Female[t]
    dM[i,j]=periode$Male.death[t]
    dK[i,j]=periode$Male.death[t]
  }
  t=t+1
}
qM_p = dM/pM
qK_p = dK/pK

#kjører over dataene og estimerer GM-parametere for periodedata

theta_mp = matrix(NA,nrow=70, ncol=3)
theta_kp = matrix(NA,nrow=70, ncol=3)
s=-c(8,9,2)

for(j in 1:70){
  thm = optim(-c(8,9,2),minusgompakll1,"BFGS",ll=(1:83),
    n=pM[,j],z=dM[,j])

```



```

        thk = optim(-c(8,9,2),minusgompakll1,"BFGS",ll=(1:83),
                    n=pK[,j],z=dM[,j])

        theta_mp[j,] = exp(thm$par)
        theta_kp[j,] = exp(thk$par)
    }

#teste om modell og parametere for periodedata
#passer rådataene for periodedata

q_x_mp = matrix(NA, nrow=(100-18+1), ncol=70)
q_x_kp = matrix(NA, nrow=(100-18+1), ncol=70)
x = (0:82)

for(i in 1:(100-18+1)){
    for(j in 1:70){
        q_x_mp[i,j] = 1-exp(-theta_mp[j,1]-theta_mp[j,2]*
                             exp(theta_mp[j,3]*x[i]))
        q_x_kp[i,j] = 1-exp(-theta_kp[j,1]-theta_kp[j,2]*
                             exp(theta_kp[j,3]*x[i]))
    }
}

```

Residual-beregninger og kohorteffekter for kohortdata

```

#Tilsvarende kode for periodedata, med noe tilpasning

#estimerer residualer for individer i alle år og aldre

res_m = matrix(NA, nrow=(100-18+1), ncol=40) #menn
res_k = matrix(NA, nrow=(100-18+1), ncol=40) #kvinner
x = (0:82)

for(i in 1:(100-18+1)){
    for(j in 1:40){
        res_m[i,j] = log(qM[i,j]/(1-qM[i,j]))-
                      (theta_menn[j,1],theta_menn[j,2]*
                       exp(theta_menn[j,3]*x[i]))
        res_k[i,j] = log(qK[i,j]/(1-qK[i,j]))-

```

```

                                (theta_kvinner[j,1]-theta_kvinner[j,2]*
                                exp(theta_kvinner[j,3]*x[i]))
                                }
                                }

#medpolish()-metoden

#menn
z = (1940:1979)
c = data.frame(z, res_m)
med_m = medpolish(c)
plot(med_m, ylab="Residualer",xlab="Diagnostisk sjekk verdier",
      panel.first=ggbg())
med_m_trend = c - med_m$residuals #gjennomsnitt
all(c == med_m$overall+outer(med_m$row,med_m$col,"+")+med_m$residuals)

#kvinner
c1 = data.frame(z, res_k)
med_k = medpolish(c1)
plot(med_k, ylab="Residualer",xlab="Diagnostisk sjekk verdier",
      panel.first=ggbg())
med_k_trend = c1 - med_k$residuals #gjennomsnitt
all(c1 == med_k$overall+outer(med_k$row,med_k$col,"+")+med_k$residuals)

#minste kvadraters metode på residualene for menn

a12k = matrix(NA, nrow=(100-18+1), ncol=40)
a21k = matrix(NA, nrow=(100-18+1), ncol=40)
a22k = matrix(NA, nrow=(100-18+1), ncol=40)
b1k = matrix(NA, nrow=(100-18+1), ncol=40)
b2k = matrix(NA, nrow=(100-18+1), ncol=40)

for(j in 1:40){
  for(i in 1:(100-18+1)){
    a12k[i,j] = aM[i,j]*(i+j)
    a21k[i,j] = aM[i,j]*(i+j)
    a22k[i,j] = aM[i,j]*(i+j)^2
    b1k = aM[i,j]*res_m[i,j]
    b2k = aM[i,j]*(i+j)
  }
}

```

```

}

a11k = sum(aM)
a12k = sum(a12k)
a21k = sum(a21k)
a22k = sum(a22k)
b1k = sum(b1k)
b2k = sum(b2k)

#2 ligninger med 2 ukjente

A_m = rbind(c(a11k,a12k),
             c(a21k,a22k))
B_m = c(b1k,b2k)

solve(A_m,B_m)

#Tilsvarende kode for kvinner

```

Programkode for dynamisk modell og fremskrivninger av GM-parametere

```

#Tilsvarende kode for periodedata, med noe tilpasning

#tilfeldig gange/random walk

#menn
ran = theta_menn[,1] #theta parameter 1
stdev = sqrt(mean(diff(ran)^2))
n.sim = 20
bar = matrix(rnorm(n.sim*length(ran),mean=0,sd=stdev),
             nrow=n.sim)
plot(seq(1,length(ran)),ran,type="o",pch=21, ylim=c(-max(rowSums(bar)),
             max(rowSums(bar))), col="red")
for(ii in 1:n.sim){
  points(seq(1,length(ran)),cumsum(bar[ii,]),type="o",pch=21)
}
#tilsvarende fremgangsmåte for theta parameter 2 og 3
#tilsvarende kode for kvinner

#simulerer baner for utvikling av parametere

```

```

m = 40
epsilon_m <- theta_menn[2:m,]-theta_menn[1:(m-1),]
epsilon_k <- theta_kvinner[2:m,]-theta_kvinner[1:(m-1),]

cholesky_m <- t(chol(cov(epsilon_m))) #menn
cholesky_k <- t(chol(cov(epsilon_k))) #kvinner

meaneps_m <- apply(epsilon_m, 2, mean)
meaneps_k <- apply(epsilon_k, 2, mean)
sigmaeps_m <- apply(epsilon_m, 2, sd)
sigmaeps_k <- apply(epsilon_k, 2, sd)

ant_arfrem <- 50
simuleringer <- 100
theta_menn.sim <- array(NA, dim=c(ant_arfrem, 3, simuleringer))
theta_kvinner.sim <- array(NA, dim=c(ant_arfrem, 3, simuleringer))
epsilon_menn.sim <- matrix(NA, nrow=ant_arfrem, ncol=3)
epsilon_kvinner.sim <- matrix(NA, nrow=ant_arfrem, ncol=3)

for(g in 1:simuleringer){
  for(k in 1:ant_arfrem){
    nu_m <- matrix(rnorm(ant_arfrem*3), nrow=ant_arfrem,
                    ncol=3)
    nu_k <- matrix(rnorm(ant_arfrem*3), nrow=ant_arfrem,
                    ncol=3)

    for(i in 1:3){
      e_m <- 0
      e_k <- 0

      for(j in 1:i){
        e_m <- e_m + cholesky_m[i,j]*nu_m[k,i]
        e_k <- e_k + cholesky_k[i,j]*nu_k[k,i]
      }
      epsilon_menn.sim[k,i] <- meaneps_m[i] + e_m
      epsilon_kvinner.sim[k,i] <- meaneps_k[i] + e_k
    }
  }
  theta_menn.sim[, ,g] <- t(t(apply(epsilon_menn.sim,2,
    cumsum))+theta_menn[m,])
  theta_kvinner.sim[, ,g] <- t(t(apply(epsilon_kvinner.sim,2,
    cumsum))+theta_kvinner[m,])
}

```

Pensjonsberegninger

```
#pensjonsalder = 67, s=1, d=0.97

#konstruerer livstabell fra kohortdødeligheter

tQx_m = rbind(1,qM)
tQx_k= rbind(1,qK)
tPx_m = array(0,c(83,40))
tPx_k= array(0,c(83,40))
tPx_m[,1] = 1
tPx_k[,1]= 1
for(x in 1:83){
  for(t in 2:40){
    tPx_m[x,t] = tPx_m[x,t-1]*(1-tQx_m[x,t-1])
    tPx_k[x,t] = tPx_k[x,t-1]*(1-tQx_k[x,t-1])
  }
}
#ved denne metoden må "onetimeprem"-funksjonen modifiseres
#til å ta inn tPx

onetimeprem=function(tPx,s,l_r,d){

  #først livstabell
  l_e=length(tPx[,lengde alder])
  kp=tPx$kp
  # så indikator matrise
  I=matrix(0,l_e+1,l_e+1)
  I[row(I)+col(I)>l_r+1]=1
  #beregner engangspremie
  ll=0:l_e
  M=s*(d**ll)*kp*I
  pi=apply(M,2,sum)
  list(pi=pi)
}

#beregner engangspremie, utvalgte år, gjennom "lifetab"-funksjonen

#menn
ep_m = onetimeprem(qM[,1],1,(67-18),0.97)$pi
ep1_m = onetimeprem(qM[,10],1,(67-18),0.97)$pi
ep2_m = onetimeprem(qM[,20],1,(67-18),0.97)$pi
```

```

ep3_m = onetimeprem(qM[,30],1,(67-18),0.97)$pi
ep4_m = onetimeprem(qM[,40],1,(67-18),0.97)$pi

#kvinner
ep_k = onetimeprem(qK[,1],1,(67-18),0.97)$pi
ep1_k = onetimeprem(qK[,10],1,(67-18),0.97)$pi
ep2_k = onetimeprem(qK[,20],1,(67-18),0.97)$pi
ep3_k = onetimeprem(qK[,30],1,(67-18),0.97)$pi
ep4_k = onetimeprem(qK[,40],1,(67-18),0.97)$pi

#engangspremie uten kohorteffekt

#fra minste kvadraters metode
a_koh_m = -1.257242e-05
b_koh_m = 2.373149e-07

a_koh_k = -2.047036e-05
b_koh_k = 3.716089e-07

premie_m = matrix(NA, nrow=(100-18+1), ncol=40)
premie_k = matrix(NA, nrow=(100-18+1), ncol=40)
x = (0:82)

for(i in 1:(100-18+1)){
  for(j in 1:40){
    premie_m[i,j] = log(qM[i,j]/(1-qM[i,j]))-
      theta_menn[j,1]-theta_menn[j,2]*
      exp(theta_menn[j,3]*x[i]-
        (a_koh_m+b_koh_m*(i+j)))
    premie_k[i,j] = log(qK[i,j]/(1-qK[i,j]))-
      theta_kvinner[j,1]-theta_kvinner[j,2]*
      exp(theta_kvinner[j,3]*x[i]-
        (a_koh_k+b_koh_k*(i+j)))
  }
}

#menn
ep_m_u = onetimeprem(premie_m[,1],1,(67-18),0.97)$pi
ep1_m_u = onetimeprem(premie_m[,10],1,(67-18),0.97)$pi
ep2_m_u = onetimeprem(premie_m[,20],1,(67-18),0.97)$pi
ep3_m_u = onetimeprem(premie_m[,30],1,(67-18),0.97)$pi
ep4_m_u = onetimeprem(premie_m[,40],1,(67-18),0.97)$pi

```

```
#kvinner
ep_k_u = onetimeprem(premie_k[,1],1,(67-18),0.97)$pi
ep1_k_u = onetimeprem(premie_k[,10],1,(67-18),0.97)$pi
ep2_k_u = onetimeprem(premie_k[,20],1,(67-18),0.97)$pi
ep3_k_u = onetimeprem(premie_k[,30],1,(67-18),0.97)$pi
ep4_k_u = onetimeprem(premie_k[,40],1,(67-18),0.97)$pi
```

8 Litteratur og kilder

- 1) Bølviken, E. (2014): *Computation and Modelling in Insurance and Finance*. Cambridge University Press, Cambridge
- 2) Mamelund og Borgan (1996): *Rapport 96/6*, Statistisk Sentralbyrå, Oslo
- 3) Human mortality database (HMD): <http://www.mortality.org/>
- 4) Andrew J. G. Cairns, David Blake, Kevin Dowd, Guy D. Coughland, David Epstein, Alen Ongd & Igor Baleviche (2009): *North American Actuarial Journal*, Volum 13, Utgave 1
- 5) <http://www.bendixcarstensen.com/APC/Carstensen.2007a.pdf>
B.Carstensen: Age-period-cohort models for the Lexis diagram
- 6) Tukey JW (MS: 1977): *Exploratory Data Analysis*. Addison-Wesley Publishing Company
- 7) Lindstrøm, T. (2006): *Kalkulus*. Universitetsforlaget
- 8) Andrew J.G. Cairns, David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein and Marwa KhalafAllah. (2011): *Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models*. The Pensions Institute, Cass Business School, United Kingdom